



TUGAS AKHIR – SM141501

**PEMROSESAN BAHASA ALAMI UNTUK
PERINGKASAN MULTI-DOKUMEN
MENGUNAKAN ALGORITMA
EKSPEKTASI MAKSIMISASI (EM)**

**ADZRA NAZIHAH
NRP 06111440000102**

**Dosen Pembimbing
Prof. Dr. Mohammad Isa Irawan, MT**

**DEPARTEMEN MATEMATIKA
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA
Institut Teknologi Sepuluh Nopember
Surabaya 2018**



FINAL PROJECT – SM141501

**NATURAL LANGUAGE PROCESSING
FOR MULTI-DOCUMENT
SUMMARIZATION USING
EXPECTATION MAXIMIZATION (EM)
ALGORITHM**

**ADZRA NAZIHAH
NRP 06111440000102**

**Supervisor
Prof. Dr. Mohammad Isa Irawan, MT**

**MATHEMATICS DEPARTMENT
FACULTY OF MATHEMATICS, COMPUTING, AND DATA
SCIENCE
Institut Teknologi Sepuluh Nopember
Surabaya 2018**

LEMBAR PENGESAHAN

PEMROSESAN BAHASA ALAMI UNTUK PERINGKASAN MULTI-DOKUMEN MENGGUNAKAN ALGORITMA EKSPEKTASI MAKSIMISASI (EM)

NATURAL LANGUAGE PROCESSING FOR MULTI- DOCUMENT SUMMARIZATION USING EXPECTATION MAXIMIZATION (EM) ALGORITHM

TUGAS AKHIR

Diajukan untuk memenuhi salah satu syarat
Untuk memperoleh gelar Sarjana Sains
Pada bidang studi Ilmu Komputer
Program Studi S-1 Departemen Matematika
Fakultas Matematika, Komputasi, dan Sains Data
Institut Teknologi Sepuluh Nopember Surabaya

Oleh :

Adzra Nazihah
NRP. 06111440000102

Menyetujui,
Dosen Pembimbing,


Prof. Dr. Mohammad Isa Irawan, MT.

NIP. 19631225 198903 1 001

Mengetahui,
Kepala Departemen Matematika
FMKSD ITS


Dr. Imam Mukhlash, S.Si, MT

NIP. 19700831 199403 1 003

Surabaya, Juli 2018

**PEMROSESAN BAHASA ALAMI UNTUK
PERINGKASAN
MULTI-DOKUMEN MENGGUNAKAN ALGORITMA
EKSPEKTASI MAKSIMISASI (EM)**

Nama Mahasiswa : Adzra Nazihah
NRP : 06111440000102
Departemen : Matematika
Dosen Pembimbing : Prof. Dr. M. Isa Irawan, MT

Abstrak

Seiring dengan berkembangnya teknologi, masyarakat disajikan dengan banyaknya informasi yang tersedia pada internet. Namun kesibukan di jaman modern ini mempersingkat waktu dan kesempatan masyarakat untuk dapat mengakses seluruh informasi yang ada dalam suatu dokumen. Kebutuhan akan sebuah sistem peringkasan yang mampu mengambil informasi penting dari beberapa dokumen menjadi tidak terelakkan. Tugas akhir ini bertujuan untuk membuat sistem peringkasan multi-dokumen dengan metode ekstraktif menggunakan algoritma EM. Peringkasan ekstraktif dilakukan dengan memilih kalimat-kalimat penting dari tiap dokumen yang kemudian diletakkan pada hasil rangkuman. Hasil dari tugas akhir ini adalah sebuah program komputer yang mengaplikasikan algoritma EM untuk peringkasan multi-dokumen. 90% rangkuman yang dihasilkan mampu merepresentasikan keseluruhan dokumen dengan bentuk yang lebih sederhana. Pengguna akan dengan mudah mendapatkan informasi dari tiap dokumen dengan hanya membaca hasil ringkasan.

Kata kunci: Peringkasan Multi-Dokumen, Peringkasan Ekstraktif, Pemrosesan Bahasa Alami, Ekspektasi Maksimisasi.

NATURAL LANGUAGE PROCESSING FOR MULTI-DOCUMENT SUMMARIZATION USING EXPECTATION MAXIMIZATION (EM) ALGORITHM

Name of Student : Adzra Nazihah
NRP : 06111440000102
Department : Mathematics
Supervisor : Prof. Dr. M. Isa Irawan, MT

Abstract

As the technology develops, people are served with many informations on the internet. However, nowadays the mass don't have time and chance to obtain all those informations. The need for summarizer capable to extract significant information becomes intolerable. This final project aims to do multi-document summarization with extractive method using EM algorithm. Extractive summarization was done by choosing important sentences from documents to be put as the summary result. The result of this project was a computer program which implemented EM algorithm to summarize information from multiple documents. 90% summary result was capable to represent a whole documents in a simple form. Users would easily acquire information only by reading the summary result.

Keywords: Multi-Document Summarization, Extractive Summarization, Natural Language Processing, Expectation Maximization.

KATA PENGANTAR

Segala puji syukur penulis panjatkan ke hadirat Allah SWT, karena dengan ridho-Nya penulis dapat menyelesaikan Tugas Akhir dengan judul

**“PEMROSESAN BAHASA ALAMI UNTUK
PERINGKASAN**

**MULTI-DOKUMEN MENGGUNAKAN ALGORITMA
EKSPEKTASI MAKSIMISASI (EM)”**

yang merupakan salah satu persyaratan akademis dalam menyelesaikan Program Sarjana Matematika, Fakultas Matematika, Komputasi, dan Sains Data, Institut Teknologi Sepuluh Nopember, Surabaya.

Tugas Akhir ini dapat diselesaikan dengan baik berkat kerja sama, bantuan, dan dukungan dari banyak pihak. Sehubungan dengan hal tersebut, penulis ingin mengucapkan terima kasih kepada:

1. Bpk. Dr. Imam Mukhlash, S.Si, MT selaku Kepala Departemen Matematika ITS.
2. Bpk. Dr. Darmaji, S.Si, MT selaku Dosen Wali yang telah memberikan arahan akademik selama penulis menempuh pendidikan di Departemen Matematika ITS.
3. Prof. Dr. Mohammad Isa Irawan, MT selaku Dosen Pembimbing yang telah membimbing penulis dalam mengerjakan Tugas Akhir ini sehingga dapat terselesaikan dengan baik.
4. Dr. Didik Khusnul Arif, S.Si, M.Si selaku Ketua Program Studi S1 Departemen Matematika ITS.
5. Drs. Iis Herisman, M.Si selaku Sekretaris Program Studi S1 Departemen Matematika ITS.
6. Seluruh jajaran dosen dan staf Departemen Matematika ITS.
7. Keluarga tercinta yang senantiasa memberikan dukungan dan do'a yang tak terhingga.

8. Muhammad Asaduddin, selaku *partner* penulis yang selalu siap membantu dan mengarahkan penulis selama proses pembuatan Tugas Akhir ini.
9. Teman-teman angkatan 2014.
10. Semua pihak yang tidak bisa penulis sebutkan satu-persatu, terima kasih telah membantu sampai terselesaikannya Tugas Akhir ini.

Penulis menyadari bahwa Tugas Akhir ini masih jauh dari kesempurnaan. Oleh karena itu, penulis mengharapkan kritik dan saran dari pembaca. Akhir kata, semoga Tugas Akhir ini dapat bermanfaat bagi semua pihak yang berkepentingan.

Surabaya, Mei 2018

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PENGESAHAN	v
KATA PENGANTAR	xi
DAFTAR ISI	xiii
DAFTAR GAMBAR	xv
DAFTAR TABEL	xvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan	3
1.5 Manfaat	3
1.6 Sistematika Penulisan Tugas Akhir	3
BAB II TINJAUAN PUSTAKA	5
2.1 Penelitian Terdahulu	5
2.2 Peringkasan Multi-Dokumen	6
2.3 Pemrosesan Bahasa Alami	8
2.4 Model Ruang Vektor	11
2.5 Algoritma Ekspektasi Maksimisasi	12
BAB III METODE PENELITIAN	17
3.1 Objek dan Aspek Penelitian	17
3.2 Peralatan	17
3.3 Tahapan Penelitian	17
BAB IV PERANCANGAN IMPLEMENTASI	23
4.1 Pemrosesan Bahasa Alami untuk Data Masukan	23
4.2 Perancangan Matriks Jarak	29
4.3 Perumusan Algoritma Ekspektasi Maksimisasi	31
4.4 Perancangan Graf Dokumen	43
BAB V IMPLEMENTASI DAN PENGUJIAN	45
5.1 Implementasi Perangkat Lunak	45
5.1.1 Tampilan Aplikasi	45

5.1.2	Pengujian Perangkat Lunak.....	51
5.2	Pengujian Perangkat Lunak.....	53
5.2.1	Summarize with String.....	53
5.2.2	Summarize with Links.....	70
5.2.3	Validasi Hasil Ringkasan	76
BAB VI PENUTUP.....		81
6.1	Kesimpulan.....	81
6.2	Saran.....	81
DAFTAR PUSTAKA.....		83
LAMPIRAN A		87
BIODATA PENULIS.....		119

DAFTAR GAMBAR

Gambar 1.1 Total Jumlah Website	1
Gambar 2.1 Data set untuk clustering menggunakan EM	14
Gambar 3.1 Diagram Alir Metode Penelitian.....	22
Gambar 4.1 Diagram Alir Algoritma EM untuk Peringkasan Multi-Dokumen	42
Gambar 5.1 Halaman Utama Program	46
Gambar 5.2 Jendela HowManyDocs	47
Gambar 5.3 Jendela Peringkasan 1 Dokumen Menggunakan String.....	47
Gambar 5.4 Jendela Peringkasan 1 Dokumen Menggunakan Link	48
Gambar 5.5 Jendela Peringkasan 2 Dokumen Menggunakan String.....	48
Gambar 5.6 Jendela Peringkasan 2 Dokumen Menggunakan Link	49
Gambar 5.7 Jendela Peringkasan 3 Dokumen Menggunakan String.....	49
Gambar 5.8 Jendela Peringkasan 4 Dokumen Menggunakan String.....	50
Gambar 5.9 Jendela Peringkasan 4 Dokumen Menggunakan String.....	50
Gambar 5.10 Jendela Peringkasan 4 Dokumen Menggunakan Link.....	51
Gambar 5.11 Tampilan Hasil	52

DAFTAR TABEL

Tabel 2.1 Contoh Matriks Ciri (Ruang Vektor)	11
Tabel 2.2 Hasil dari Tiga Iterasi Pertama Bobot Keanggotaan dengan Algoritma EM.....	14
Tabel 2.3 Langkah-Langkah Algoritma Ekspektasi Maksimisasi	16
Tabel 4.1 Contoh Nilai TF-IDF Kata	30
Tabel 4.2 Contoh Nilai Jarak Kata terhadap Titik Tengah Tiap Cluster	31
Tabel 4.3 Contoh Bobot Keanggotaan Kata terhadap Cluster pada Iterasi Terakhir	33
Tabel 4.4 Contoh Pencarian Titik Tengah Terbaik	35
Tabel 4.5 Contoh Hasil Clustering Kata Pada tiap Iterasi	36
Tabel 5.1 Hasil Clustering	71
Tabel 5.2 Hasil Perbandingan Jumlah Kata.....	77
Tabel 5.3 Hasil Pengujian Nilai Similaritas Intra dan Inter.....	78
Tabel A Data Clustering Kata	87
Tabel B Data Clustering Kata.....	101
Tabel C Data Clustering Kata.....	103
Tabel D Data Clustering Kata	117

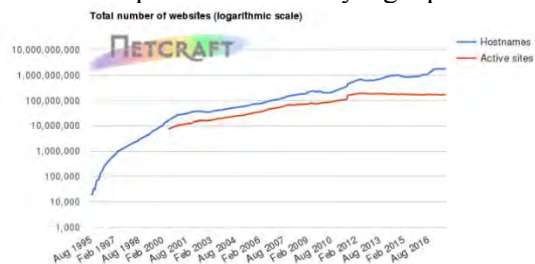
BAB I

PENDAHULUAN

Pada bab ini dibahas mengenai latar belakang yang mendasari penulisan Tugas Akhir ini. Di dalamnya mencakup identifikasi permasalahan pada topik Tugas Akhir kemudian dirumuskan menjadi permasalahan yang diberikan batasan-batasan dalam pembahasan pada Tugas Akhir ini.

1.1 Latar Belakang

Koleksi dokumen pada internet terus meningkat dalam dua dekade terakhir. Gambar 1.1 adalah grafik yang memperlihatkan kenaikan jumlah laman aktif pada internet dari tahun 1995 sampai dengan 2016. Data tersebut diambil dari website Netcraft [1]. Jaman sekarang pengaksesan informasi sering dilakukan dengan menggunakan mesin pencarian Google. Hasil pencarian informasi yang diperoleh adalah berupa kumpulan link dari banyak dokumen yang kemudian akan dipilih oleh pengguna sesuai dengan apa yang mereka inginkan. Pengguna melakukan pemilihan dokumen dengan membaca beberapa kata dari dokumen yang dihubungkan dengan link sebagai hasil dari query. Pemilihan beberapa kata dari dokumen tidak cukup untuk merepresentasikan keseluruhan query sehingga tak jarang pengguna harus mendownload dokumen tersebut untuk mendapatkan informasi yang tepat.



Gambar 1.1 Total Jumlah Website

Berdasarkan Lee [2], dengan menggunakan internet, jumlah dokumen pada domain publik terus meningkat sehingga

menyebabkan terjadinya redundansi atas banyaknya dokumen tersebut. Pengembangan dari peringkasan dokumen adalah kunci untuk menyelesaikan masalah tersebut. Hal ini dilakukan dengan mengembangkan metode otomatis untuk mengekstrak informasi yang paling berhubungan dengan teks, yang diteliti oleh *Automatic Generation of Summaries* (AGS) [3, 4, 5, 6].

Automatic Text Summarization (ATS) adalah area riset aktif yang berhubungan dengan peringkasan satu dan multi-dokumen. Pada peringkasan satu dokumen, peringkasan dilakukan dari satu dokumen saja. Sedangkan pada peringkasan multi-dokumen, peringkasan dilakukan dari koleksi dokumen. Dalam tugas akhir ini, peringkasan multi-dokumen lah yang dilakukan.

Berdasarkan Yulia [7], metode peringkasan dapat dibagi menjadi dua, abstraktif dan ekstraktif. Ringkasan abstraktif adalah sembarang teks yang mendeskripsikan konteks dari sumber dokumen. Proses ringkasan abstraktif terdiri dari mengartikan dokumen asli dan mengungkapkan kembali dokumen tersebut dalam beberapa kata. Sedangkan ringkasan ekstraktif dibentuk dengan memilih beberapa kalimat penting dari dokumen asli kemudian diberikan ke pengguna. Metode peringkasan ekstraktif hanya menentukan apakah sebuah kalimat dimasukkan ke hasil ringkasan atau tidak. Struktur kalimat dari hasil peringkasan cenderung kaku bila dibandingkan dengan metode abstraktif. Namun kesederhanaan atas teknik perhitungan membuat peringkasan ekstraktif menjadi metode peringkasan yang language-independent. Dalam tugas akhir ini, dilakukan peringkasan ekstraktif.

Penelitian peringkasan teks sudah banyak dilakukan menggunakan beberapa algoritma seperti K-means dan Ekspektasi Maksimisasi. Menurut penelitian yang dilakukan oleh Yllias Chali, Shafiq R. Jothy, dan Sadid A. Hasan, algoritma Ekspektasi Maksimisasi memberikan hasil yang lebih baik dibandingkan dengan algoritma K-Means [8]. Hal tersebut menjadi landasan bagi saya untuk memilih Ekspektasi Maksimisasi sebagai metode dalam tugas akhir yang saya kerjakan.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan, rumusan masalah yang dapat diambil adalah bagaimana mengimplementasikan algoritma EM untuk peringkasan multi-dokumen.

1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Maksimal jumlah dokumen masukan ke dalam sistem yang dibuat adalah sebanyak 4 dokumen.
2. Dokumen masukan berupa multi-dokumen berita berbahasa Inggris atau Indonesia dengan topik yang sama.
3. Kumpulan berita diambil dari *google.com*.
4. Bentuk masukan teks ke sistem berupa *string* atau *link*.
5. Pengambilan data pada pemrosesan ringkasan dokumen dapat dilakukan secara *online* maupun *offline*.
6. Peringkasan teks otomatis berbasis program python yang dapat melakukan peringkasan multi-dokumen.

1.4 Tujuan

Tujuan dari tugas akhir ini adalah untuk melakukan peringkasan multi-dokumen menggunakan algoritma Ekspektasi Maksimisasi dan memeriksa ketepatan ringkasan yang dihasilkan.

1.5 Manfaat

Manfaat dari tugas akhir ini adalah menghasilkan program yang dapat meringkas informasi dari beberapa dokumen yang memiliki kesamaan topik.

1.6 Sistematika Penulisan Tugas Akhir

Sistematika dari penulisan Tugas Akhir ini adalah sebagai berikut :

1. BAB I PENDAHULUAN

Bab ini menjelaskan tentang gambaran umum dari penulisan Tugas Akhir ini yang meliputi latar belakang masalah,

rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

2. **BAB II TINJAUAN PUSTAKA**

Bab ini berisi tentang materi-materi yang mendukung Tugas Akhir ini, antara lain penelitian terdahulu, Peringkasan Multi-Dokumen, Pemrosesan Bahasa Alami, Model Ruang Vektor, dan Algoritma Ekspektasi Maksimisasi.

3. **BAB III METODOLOGI PENELITIAN**

Pada bab ini dibahas tentang langkah-langkah dan metode yang digunakan untuk menyelesaikan Tugas Akhir ini.

4. **BAB IV PERANCANGAN IMPLEMENTASI**

Pada bab ini akan diuraikan bagaimana tahapan-tahapan dalam perancangan implementasi.

5. **BAB V IMPLEMENTASI**

Bab ini menjelaskan mengenai implementasi dari algoritma klustering Ekspektasi Maksimisasi untuk meringkas dokumen-dokumen yang memiliki kesamaan topik. Setelah itu, dilakukan analisis terhadap hasil implementasi. Analisis ini bertujuan untuk melihat apakah algoritma tersebut dapat diterapkan untuk melakukan peringkasan multi-dokumen.

6. **BAB VI PENUTUP**

Bab ini berisi kesimpulan yang diperoleh dari pembahasan masalah sebelumnya serta saran yang diberikan untuk pengembangan selanjutnya.

BAB II

TINJAUAN PUSTAKA

Pada bab ini dibahas mengenai dasar teori yang digunakan dalam penyusunan Tugas Akhir ini. Dasar teori yang dijelaskan dibagi menjadi beberapa subbab yaitu Penelitian Terdahulu, Peringkasan Multi-Dokumen, Pemrosesan Bahasa Alami, Model Ruang Vektor, dan Algoritma Ekspektasi Maksimisasi.

2.1 Penelitian Terdahulu

Penelitian terdahulu yang berkaitan dengan tugas akhir yang akan saya kerjakan adalah penelitian yang dilakukan oleh Yllias Chali, dkk pada tahun 2009 mengenai Complex Question Answering. Penelitian yang dilakukan adalah tentang perhitungan kepentingan relatif atas setiap kalimat menggunakan metode empirik, K-means, dan Ekspektasi Maksimisasi. Hasil dari perhitungan kepentingan relatif dari tiga metode tersebut kemudian dibandingkan dan hasilnya adalah bahwa metode empirik menghasilkan perhitungan relatif yang lebih baik dibandingkan dengan dua metode lainnya, dan metode EM lebih baik dibandingkan dengan metode K-Means. Dalam penelitian tersebut, pengestrakan fitur-fitur yang berbeda (leksikal, semantik leksikal, kesamaan cosine, elemen basik, dan kernel berbasis sintaktik, dan shallow-semantik) dilakukan untuk mengukur kepentingan dan relevansi atas tiap kalimat. [8]

Pada tahun 2012, Meghana. N. Ingole, dkk melakukan penelitian tentang peringkasan teks. Peringkasan teks dilakukan menggunakan algoritma Ekspektasi Maksimisasi (EM). Penelitian tersebut mengimplementasikan algoritma clustering EM untuk menemukan kesamaan antar kalimat yang kemudian dapat digunakan untuk meringkas teks. Penelitian dilakukan dengan beberapa proses yaitu, (1) Input, (2) NLP Parser Engine, (3) Build Distance Matrix, (4) User fires a query, (5) Clustering. Tugas akhir yang akan saya lakukan memiliki kesamaan proses dengan penelitian tersebut namun dengan jenis inputan file yang berbeda. Penelitian tersebut hanya meringkas satu dokumen, sedangkan

tugas akhir yang akan saya lakukan adalah meringkas banyak dokumen (Multi-Document Summarization). [9]

Pada tahun 2017, Ziqiang Cao, dkk melakukan penelitian mengenai peringkasan multi-dokumen. Dalam penelitian tersebut, diajukan sebuah sistem baru bernama TCSum yang merupakan sistem peringkasan multi-dokumen menggunakan klasifikasi teks. Klasifikasi teks menghasilkan representasi teks yang lebih baik dibandingkan dengan metode peringkasan. Representasi teks ini kemudian dipakai untuk melatih model peringkasan yang lebih efektif. Terlebih lagi, dalam penelitian tersebut disebutkan jika kategori dari sebuah dokumen diketahui, maka akan tersedia kesempatan untuk menyelidiki lebih banyak tipe ringkasan. TCSum memproyeksikan setiap dokumen ke dalam sebuah representasi terdistribusi yang dibagi menjadi dua tugas. Untuk klasifikasi teks, embedding dokumen diikuti klasifikasi pemisahan untuk mempelajari hubungannya dengan kategori. Untuk peringkasan, embedding dokumen diubah untuk menyesuaikan artinya dengan referensi ringkasan. [10]

2.2 Peringkasan Multi-Dokumen

Pengolahan teks otomatis adalah bidang penelitian yang saat ini sangat aktif. Satu tugas penting di bidang ini adalah peringkasan otomatis, yang terdiri dari pengurangan ukuran teks dan penyimpanan isi informasinya [4, 11]. Summarizer adalah sistem yang menghasilkan representasi ringkas dari teks yang dimaksudkan untuk konsumsi pengguna [12].

Proses peringkasan otomatis dapat dibagi menjadi tiga tahap [11]: (1) Tahap preprocessing: diperoleh representasi terstruktur dari teks asli; (2) Tahap pemrosesan: sebuah algoritma harus mengubah struktur teks menjadi sebuah ringkasan struktur; dan (3) Tahap generasi: rangkuman akhir diperoleh dari struktur ringkasan.

Metode peringkasan dapat diklasifikasikan berdasarkan ruang linguistiknya, dalam dua kelompok besar [12]: (a) pendekatan dangkal, yaitu terbatas pada tingkat representasi sintaksis dan mencoba mengekstrak bagian yang menonjol dari teks dengan cara yang mudah; dan (b) pendekatan yang lebih

dalam, yang mengasumsikan tingkat semantik representasi teks asli dan melibatkan pemrosesan linguistik pada tingkat tertentu.

Pada pendekatan pertama, tujuan dari tahap preprocessing adalah mengurangi dimensi ruang representasi, dan biasanya mencakup: (i) eliminasi stop-word atau kata-kata umum tanpa semantik yang informasinya tidak relevan untuk tugas tersebut (mis., "the", "a") dihapuskan; (ii) tingkatan kasus: terdiri dari mengubah semua karakter menjadi huruf yang sama atau baik huruf besar atau huruf kecil; (iii) stemming: kata-kata sintaktis serupa, seperti bentuk jamak, variasi verbal, dll dianggap serupa; Tujuan dari prosedur ini adalah untuk mendapatkan asal kata atau dasar setiap kata, yang menekankan semantiknya.

Model teks yang sering digunakan adalah model vektorial [13]. Setelah langkah preprocessing setiap elemen teks (sebuah kalimat dalam kasus peringkasan teks) dianggap sebagai vektor dimensi N . Hal ini memungkinkan untuk menggunakan beberapa metrik ruang vektor tersebut untuk mengukur kesamaan antara elemen teks.

Evaluasi kualitas ringkasan yang dihasilkan adalah titik kunci dalam penelitian evaluasi peringkasan tentang peringkasan dilakukan di TIPSTER Text Summarization Evaluation Conference (SUMMAC) [14], sebagai bagian dari usaha untuk menstandarisasi prosedur uji perangkuman. Dalam hal ini, rujukan koleksi ringkasan disediakan oleh manusia agar memungkinkan perbandingan langsung atas kinerja sistem yang berpartisipasi dalam konferensi. Namun, manusia memerlukan usaha yang besar untuk membuat ringkasan semacam itu. Masalah lain yang dilaporkan adalah terkait dengan hasil ringkasan yang dilakukan antar manusia memiliki similaritas yang rendah, hanya 46% menurut Mitra [15]. Terlebih lagi: ringkasan yang dihasilkan oleh manusia tersebut pada tanggal yang berbeda memiliki kesepakatan hanya 55% [16].

Berdasarkan Joel [17], gagasan rujukan referensi yang dilakukan oleh manusia itu penting, karena jika dipertimbangkan secara objektif, dapat dilakukan evaluasi kinerja rangkuman

otomatis prosedur generasi menggunakan presisi Informasi Retrieval (IR) klasik dan mengingat ukurannya. Pada hal ini, kalimat disebut benar jika tergolong dalam rujukan ringkasan. Presisi adalah rasio jumlah kalimat terpilih yang benar terhadap jumlah total kalimat terpilih, dan recall adalah rasio dari jumlah kalimat terpilih yang benar terhadap jumlah total kalimat yang benar. Dalam kasus perangkuman panjang, kedua ukuran itu identik, karena ukuran rujukan dan ringkasan ekstraktif yang diperoleh secara otomatis adalah identik.

Mani dan Bloedorn [18] mengusulkan sebuah prosedur otomatis untuk menghasilkan rujukan ringkasan: jika setiap teks asli berisi ringkasan yang diberikan oleh penulis, rujukan ringkasan ekstraksi ukuran yang sesuai terdiri dari K yang paling mirip dengan kalimat ringkasan yang penulis berikan, sesuai dengan ukuran kosinus. Menggunakan Pendekatan ini, mudah untuk mendapatkan referensi ringkasan, bahkan untuk koleksi dokumen besar.

2.3 Pemrosesan Bahasa Alami

Menurut Ronan [19], sebuah program komputer tidak bisa mengubah teks Inggris menjadi bentuk struktur data yang ramah bagi programmer dan tidak bisa menggambarkan makna teks bahasa alami. Belum ada konsensus tentang bentuk atau keberadaan struktur data semacam itu. Sampai masalah dasar Artificial Intelligence tersebut dipecahkan, ilmuwan komputer harus menyelesaikan masalah tentang ekstraksi representasi sederhana yang menggambarkan aspek terbatas dari informasi tekstual. Representasi sederhana ini dimotivasi oleh aplikasi spesifik (misalnya, varian kata-kata untuk pencarian informasi), atau dengan keyakinan bahwa komputer menangkap sesuatu yang lebih umum tentang bahasa alami. Komputer dapat menjelaskan informasi sintaksis (misalnya, part-of-speech tagging, chunking, and parsing) atau informasi semantik (misalnya, merasakan disambiguasi kata, peran pelabelan semantik, ekstraksi entitas, dan resolusi anafora). Teks corpora telah dijelaskan secara manual

dengan struktur data tersebut agar bisa membandingkan kinerja berbagai sistem. Ketersediaan tolak ukur standar telah mendorong penelitian Natural Language Processing (NLP) dan sistem yang efektif telah dirancang untuk semua tugas ini. Sistem seperti ini sering dipandang sebagai komponen perangkat lunak untuk membangun solusi pemrosesan bahasa alami pada dunia nyata.

Pemrosesan bahasa alami mencakup banyak teknik berbeda untuk menafsirkan bahasa manusia, mulai dari metode statistik dan pembelajaran mesin hingga pendekatan berbasis aturan dan algoritmik. Tugas NLP dasar termasuk pemisahan kalimat, tokenisasi, lemmatization / stemming, tagging part-of-speech, stopwords removal, deteksi bahasa, dan identifikasi hubungan semantik. [20]

Pemisahan kalimat merupakan proses pemecahan dokumen menjadi kalimat-kalimat. Tokenisasi adalah proses pemisahan dokumen menjadi token-token atau bagian-bagian tertentu. Pada umumnya setiap kata teridentifikasi atau terpisahkan dengan kata yang lain oleh karakter spasi, sehingga proses tokenisasi mengandalkan karakter spasi pada dokumen untuk melakukan pemisahan kata. Stemming merupakan salah satu proses dari pembuatan sistem temu kembali, dimana proses stemming akan dilakukan setelah proses filtering. Proses stemming ini membuat term yang ada pada tabel filtering menjadi kata dasar, dengan menghilangkan semua imbuhan yang ada pada kata tersebut (imbuhan meng-, me-, kan-, di- , i, pe, peng-, a-, dll.). POS (Part-of-Speech) Tag merupakan suatu cara pengkategorian kelas kata, seperti kata benda, kata kerja, kata sifat, dll. POS Tagger merupakan sebuah aplikasi yang mampu melakukan proses anotasi part-of-speech tag untuk setiap kata di dalam dokumen secara otomatis. *Stopwords* adalah kata umum (common words) yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna. Stop words umumnya dimanfaatkan dalam task information retrieval. Contoh stop words untuk bahasa Inggris diantaranya “of”, “the”. Sedangkan untuk bahasa Indonesia

diantaranya “yang”, “di”, “ke”. *Stopwords Removal* adalah proses penghapusan kata umum tersebut.

Secara umum, tugas NLP memecah bahasa menjadi potongan-potongan elemen yang lebih pendek, mencoba memahami hubungan di antara potongan-potongan dan menjelajahi bagaimana potongan-potongan itu bekerja bersama untuk menciptakan makna. Tugas-tugas yang mendasarinya sering digunakan dalam kemampuan NLP tingkat yang lebih tinggi, seperti:

- Kategorisasi konten. Ringkasan dokumen berbasis linguistik, termasuk pencarian dan pengindeksan, pemberitahuan konten dan deteksi duplikasi.
- Penemuan dan pemodelan topik. Tangkap makna dan tema secara akurat dalam koleksi teks, dan terapkan analitik tingkat lanjut ke teks, seperti pengoptimalan dan perkiraan.
- Ekstraksi kontekstual. Secara otomatis menarik informasi terstruktur dari sumber berbasis teks.
- Analisis sentimen. Mengidentifikasi suasana hati atau opini subjektif dalam jumlah besar teks, termasuk sentimen rata-rata dan penambangan opini.
- Konversi ucapan-ke-teks dan teks-ke-ucapan. Mentransmisikan perintah suara menjadi teks tertulis, dan sebaliknya.
- Rangkuman dokumen. Secara otomatis menghasilkan sinopsis dari kumpulan besar teks.
- Mesin penerjemah. Terjemahan teks atau ucapan otomatis dari satu bahasa ke bahasa lainnya.

Dalam semua kasus ini, tujuan utamanya adalah untuk mengambil masukan bahasa mentah dan menggunakan linguistik dan algoritma untuk mengubah atau memperkaya teks sedemikian rupa sehingga memberikan nilai yang lebih besar [20].

2.4 Model Ruang Vektor

Untuk dapat mengelompokkan dokumen, kumpulan dokumen tersebut harus dikonversi terlebih dahulu menjadi sebuah ruang vektor. Ruang vektor ini berupa matriks dua dimensi dan sering disebut dengan Vector Space Model (VSM) [21]. Tabel 2.1 merupakan contoh VSM, dengan baris menyatakan dokumen sedangkan kolom menyatakan ciri yang terkandung. Oleh karena itu, jumlah baris VSM sama dengan jumlah data yang diobservasi. Jenis ciri yang digunakan adalah ciri frekuensi kemunculan kata. Mengingat besarnya dimensi ciri, maka frekuensi kata yang digunakan sebagai dasar pembuatan matriks ciri adalah katakata yang memiliki kemunculan lebih dari sepuluh kali. Katakata yang kemunculannya berada di bawah batas yang ditentukan dianggap belum bisa mewakili makna suatu dokumen.

Tabel 2.1 Contoh Matriks Ciri (Ruang Vektor)

	Ciri 1	Ciri 2	Ciri 3	...	Ciri N
Dok 1	$w_{1,1}$	$w_{2,1}$	$w_{3,1}$...	$w_{n,1}$
Dok 2	$w_{1,2}$	$w_{2,2}$	$w_{3,2}$...	$w_{n,2}$
Dok 3	$w_{1,3}$	$w_{2,3}$	$w_{3,3}$...	$w_{n,3}$

TF/IDF merupakan mekanisme pembobotan suatu ciri yang berdasar pada frekuensi kemunculan kata. Secara sederhana, bobot TF/IDF memberikan nilai yang tinggi untuk kata yang sering muncul di satu dokumen tetapi jarang muncul di dokumen lainnya, serta nilai yang rendah untuk kata yang berfrekuensi tinggi pada suatu dokumen tetapi juga sering muncul di dokumen lain [22]. Persamaan bobot TF/IDF dituliskan pada persamaan berikut.

$$tf - idf_{t,d} = tf_{t,d} \cdot idf_t \quad (2.1)$$

dengan $tf - idf_{t,d}$ merupakan bobot TF/IDF kata_t pada dokumen_d, tf merupakan frekuensi kata_t pada dokumen_d, dan idf_t merupakan nilai idf dari kata_t, dengan

$$idf_{t,d} = \log \frac{N}{df_t} \quad (2.2)$$

N merupakan jumlah dokumen yang diproses dan df_t adalah jumlah dokumen yang mengandung kata _{t} .

Matriks jarak digunakan sebagai masukan untuk pengelompokan algoritma yang memberikan perbedaan kalimat secara asosiatif. Perancangan matriks jarak dilakukan untuk mengetahui kesamaan dari objek masukan dengan menghitung jarak *Euclidean* antar objek. Jarak *Euclidean* menghitung jarak dua buah point dengan mengetahui nilai dari masing-masing atribut pada kedua point tersebut. Berikut formula yang digunakan untuk menghitung jarak dengan *Euclidean distance*:

$$dist(\hat{p}, \hat{q}) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (2.3)$$

dimana \hat{p} dan \hat{q} adalah vektor *Euclidean* [23].

2.5 Algoritma Ekspektasi Maksimisasi

Klaster secara umum digunakan untuk melihat kesamaan antar objek dan mengelompokkannya ke dalam beberapa bagian. Peran klastering dalam ilmu pengetahuan menjadi semakin pesat berkembang seiring dengan kebutuhannya yang semakin meningkat, diantaranya berupa bioinformatika, bisnis dan pemasaran, pendidikan, dan pemerintahan [24]. Algoritma EM adalah salah satu algoritma *clustering* yang memiliki prosedur iteratif dari dua tahap bolak-balik yaitu sebuah tahap ekspektasi yang diikuti dengan langkah maksimisasi. Tahap ekspektasi berkaitan dengan variabel dasar yang tidak diketahui, menggunakan perkiraan parameter dan kondisi pada saat pengamatan. Tahap maksimisasi memberikan perkiraan parameter baru. Pada setiap iterasi, parameter yang diestimasi memberikan peningkatan fungsi *maximum-likelihood* (ML) sampai lokal

maksimum tercapai. Inisialisasi k distribusi parameter, setiap parameter distribusi bersesuaian dengan pusat kluster. Iterasikan dua langkah maksimal umum dari fungsi *likelihood*. Hal ini tergantung pada titik-titik awal.

Langkah-langkah algoritma EM adalah sebagai berikut:

a. *Guess Model Parameter*

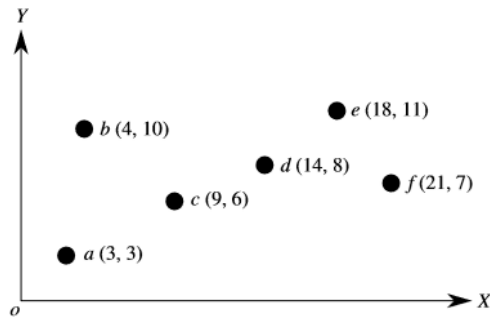
Proses ini adalah melakukan penentuan nilai tengah (*center point*) atas masing-masing kluster. Pemilihan nilai tengah dilakukan secara acak. Misalkan $c_1 = a$ dan $c_2 = b$, sebagai pusat awal dari dua kelompok. Iterasi pertama melakukan langkah ekspektasi dan langkah maksimisasi sebagai berikut [25].

b. Tahap Ekspektasi

Dalam langkah ekspektasi, untuk setiap poin dihitung derajat keanggotaannya di setiap kluster. Untuk sembarang titik o , o akan ditetapkan terhadap c_1 dan c_2 dengan bobot keanggotaan masing-masing, di mana $dist(,)$ adalah jarak Euclidean yang ada pada persamaan (2.3). Pada dasarnya, jika o dekat dengan c_1 dan $dist(o, c_1)$ kecil, tingkat keanggotaan o dengan c_1 harus tinggi. Derajat keanggotaan juga akan dinormalkan sehingga jumlah derajat untuk suatu objek sama dengan 1.

$$w_{o,c_1} = \frac{\frac{1}{(dist(o, c_1))^2}}{\frac{1}{(dist(o, c_1))^2} + \frac{1}{(dist(o, c_2))^2}} = \frac{(dist(o, c_2))^2}{(dist(o, c_1))^2 + (dist(o, c_2))^2} \quad (2.4)$$

$$w_{o,c_2} = \frac{\frac{1}{(dist(o, c_2))^2}}{\frac{1}{(dist(o, c_1))^2} + \frac{1}{(dist(o, c_2))^2}} = \frac{(dist(o, c_1))^2}{(dist(o, c_1))^2 + (dist(o, c_2))^2} \quad (2.5)$$



Gambar 2.1 Data set untuk clustering menggunakan EM

Tabel 2.2 Hasil dari Tiga Iterasi Pertama Bobot Keanggotaan dengan Algoritma EM

<i>Iteration</i>	<i>E-Step</i>	<i>M-Step</i>
1	$M^T = \begin{bmatrix} 1 & 0 & 0.48 & 0.42 & 0.41 & 0.47 \\ 0 & 1 & 0.52 & 0.58 & 0.59 & 0.53 \end{bmatrix}$	$c_1 = (8.47, 5.12)$ $c_2 = (10.42, 8.99)$
2	$M^T = \begin{bmatrix} 0.73 & 0.49 & 0.91 & 0.26 & 0.33 & 0.42 \\ 0.27 & 0.51 & 0.09 & 0.74 & 0.67 & 0.58 \end{bmatrix}$	$c_1 = (8.51, 6.11)$ $c_2 = (14.42, 8.69)$
3	$M^T = \begin{bmatrix} 0.80 & 0.76 & 0.99 & 0.02 & 0.14 & 0.23 \\ 0.20 & 0.24 & 0.01 & 0.98 & 0.86 & 0.77 \end{bmatrix}$	$c_1 = (6.40, 6.24)$ $c_2 = (16.55, 8.64)$

Pada Tabel 2.2 terlihat bahwa untuk titik a, $w_{a,c_1} = 1$ dan $w_{a,c_2} = 0$. Karena $w_{a,c_1} > w_{a,c_2}$ maka a secara eksklusif termasuk ke dalam c_1 . Untuk titik b, $w_{b,c_1} = 0$ dan $w_{b,c_2} = 1$. Untuk titik c, $w_{c,c_1} = \frac{41}{45 + 41} = 0,48$ dan $w_{c,c_2} = \frac{45}{45 + 41} = 0,52$. Derajat keanggotaan dari titik-titik lainnya ditunjukkan dalam matriks partisi pada Tabel 2.2

c. Tahap Maksimisasi

Dalam tahap maksimisasi, centroid dihitung ulang sesuai dengan matriks partisi, meminimalkan SSE yang diberikan dalam Persamaan. (11.4). Centroid baru harus disesuaikan

$$c_j = \frac{\sum_{each\ point\ o} w_{o,c_j}^2 o}{\sum_{each\ point\ o} w_{o,c_j}^2} \quad (2.6)$$

dimana $j = 1,2$.

Dilakukan pengulangan iterasi, dimana setiap iterasi berisi tahap ekspektasi dan maksimisasi.

Tabel 2.3 Langkah-Langkah Algoritma Ekspektasi Maksimisasi

Step	
1	Inisialisasi titik tengah atas 2 cluster secara acak $c_1 = a$ dan $c_2 = b$
2	Hitung bobot keanggotaan dari masing-masing kata terhadap 2 cluster yang ada $w_{o,c_1} = \frac{(dist(o, c_2))^2}{(dist(o, c_1))^2 + (dist(o, c_1))^2}$ $w_{o,c_2} = \frac{(dist(o, c_1))^2}{(dist(o, c_1))^2 + (dist(o, c_1))^2}$
3	Hitung titik tengah baru $c_j = \frac{\sum_{each\ point\ o} w_{o,c_j}^2 o}{\sum_{each\ point\ o} w_{o,c_j}^2}$
4	Ulangi langkah 2 dan 3 sampai $\sum_{i=0}^n a_i - a_{bri} \leq 0.000001$ $\sum_{i=0}^n b_i - b_{bri} \leq 0.000001$ <p>Dimana a_{br} dan b_{br} adalah titik tengah baru yang dihasilkan pada tahap maksimisasi sedangkan a dan b adalah titik tengah yang digunakan pada tahap ekspektasi.</p>

BAB III METODE PENELITIAN

Pada bab ini dijelaskan objek, peralatan, dan langkah-langkah yang digunakan dalam penyusunan Tugas Akhir. Disamping itu, dijelaskan pula prosedur dan proses pelaksanaan setiap langkah yang dilakukan dalam menyelesaikan Tugas Akhir.

3.1 Objek dan Aspek Penelitian

Objek yang digunakan dalam penelitian ini adalah multi-dokumen yang berbentuk teks artikel berita. Sedangkan aspek penelitian yang diteliti adalah menyelesaikan peringkasan multi-dokumen dengan algoritma EM.

3.2 Peralatan

Penelitian ini menggunakan peralatan berupa perangkat keras dan perangkat lunak, antara lain:

1. Perangkat keras

Perangkat keras yang digunakan yaitu personal komputer atau laptop dengan spesifikasi:

- Sistem operasi windows 10, 64-bit.
- Prosesor Intel(R) Core(TM) i5-7200U CPU @ 2.50GHz.
- RAM 4.00 GB.

2. Perangkat Lunak

Perangkat lunak yang digunakan adalah Anaconda dengan aplikasi Jupyter Notebook 5.0.0 dan menggunakan bahasa pemrograman python 3.6.

3.3 Tahapan Penelitian

Langkah-langkah yang digunakan dalam menyelesaikan Tugas Akhir ini yang diringkas dalam Gambar 3.1 adalah sebagai berikut :

1. Studi Literatur

Studi Literatur ini dilakukan untuk identifikasi permasalahan dengan mencari referensi yang menunjang penelitian yang berupa Tugas Akhir, jurnal internasional, buku, maupun artikel yang berhubungan dengan topik Tugas Akhir ini.

2. Implementasi

Implementasi pada penelitian ini dibuat dengan bahasa pemrograman Python 3 dan menggunakan perangkat lunak Anaconda dan aplikasi Jupyter Notebook 5.0.0. Pada tahap ini dilakukan proses dimana persoalan diterjemahkan melalui suatu rangkaian aktivitas sesuai model proses, metode, dan alat bantu yang digunakan. Berikut adalah tahapan dari proses implementasi:

a. Pemrosesan Bahasa Alami atas Data Masukan

Dalam Tugas Akhir yang saya kerjakan, Pemrosesan Bahasa Alami dilakukan atas data masukan berupa koleksi berita berbahasa inggris. Pemrosesan Bahasa Alami adalah *preprocessing* yang dilakukan dengan tahap sebagai berikut:

(1) *Split Sentence*

Mengenali akhir sebuah kalimat bukanlah tugas yang mudah untuk komputer. Data masukan dibagi menjadi kalimat terpisah dengan karakter baris baru dan diubah menjadi array paragraf dengan menggunakan metode split. Ini bisa dilakukan dengan memperlakukan masing-masing karakter '.', '!', '?' sebagai pemisah daripada penanda akhir kalimat yang pasti.

(2) Tokenisasi

Tahap ini akan memisahkan input teks menjadi token yang terpisah. Teks bisa dipisahkan menjadi token. Tanda baca, spasi dan kata terminator adalah kata melanggar karakter.

(3) *Stopword Removal*

Tahapan untuk menghilangkan stopwords disebut dengan Stopword Removal. Stopword atau kata yang tidak berguna dalam suatu kalimat terkadang bisa diabaikan. Tanpa keberadaan stopwords, suatu kalimat masih memiliki makna yang utuh.

- b. Perancangan Matriks Jarak
 - (1) Mencari Nilai TF-IDF pada persamaan (2.1) dari kata pada tiap dokumen.
 - (2) Mencari jarak antara tiap kata dengan titik tengah cluster menggunakan rumus Jarak Euclidean pada Persamaan (2.3).
- c. Perumusan Algoritma Ekspektasi Maksimisasi
 Algoritma EM adalah algoritma yang terdiri dari 2 tahap yaitu tahap ekspektasi dan maksimisasi. Pada algoritma ini parameter utamanya adalah titik a dan b untuk mendapatkan nilai w_{o,c_j} yang kemudian digunakan untuk keanggotaan cluster.

Secara singkat, langkah-langkah pengelompokan dokumen dengan algoritma *Unsupervised* EM adalah:

- 1. *Guess Initial Model Parameter*
- 2. Tahap ekspektasi: Menghitung bobot
- 3. Tahap maksimisasi: Mencari titik tengah baru
- 4. Ulangi langkah 2 dan 3 sampai konvergen. Nilai probabilitas klaster data bersifat konvergen jika pengupdetan probabilitas data terhadap klaster data tidak berubah-ubah lagi. Dengan kata lain nilai probabilitas dokumen terhadap sebuah klaster sudah bernilai 1.
- d. Perancangan Graf Dokumen
 Perancangan Graf Dokumen dengan memilih 5 kata dengan bobot keanggotaan tertinggi pada masing-masing cluster. Kalimat dari tiap dokumen sumber yang memuat kata-kata tersebut akan dijadikan sebagai hasil ringkasan.
- e. Validasi Hasil Ringkasan
 Akurasi hasil ringkasan diuji dengan dua cara. Pertama dilihat dari perbandingan jumlah kata pada hasil ringkasan dengan jumlah kata pada dokumen sumber. Cara ini bertujuan untuk membuktikan bahwa sistem mampu meringkas dokumen sumber menjadi lebih sederhana. Kedua dengan membandingkan nilai similaritas antar sumber dokumen dan nilai similaritas hasil ringkasan dengan tiap sumber dokumen. Cara ini bertujuan untuk

membuktikan bahwa hasil ringkasan memuat informasi yang ada pada dokumen sumber.

Jumlah kata pada hasil ringkasan dihitung dan dibandingkan dengan jumlah kata pada sumber dokumen. Apabila jumlah kata pada hasil ringkasan lebih sedikit dibandingkan dengan jumlah kata pada sumber dokumen, maka hasil ringkasan dapat dikatakan lebih sederhana.

Similaritas dihitung menggunakan *cosine similarity* [26] dengan rumus sebagai berikut,

$$\begin{aligned}
 similarity &= \cos(\theta) \\
 &= \frac{A \cdot B}{||A|| ||B||} \\
 &= \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (3.1)
 \end{aligned}$$

Berdasarkan Persamaan 3.1 jika diterapkan dalam dokumen adalah apabila nilai similaritas hasil ringkasan dengan tiap sumber dokumen lebih tinggi dibandingkan dengan nilai similaritas antar sumber dokumen, hasil ringkasan dapat dikatakan valid. Hasil ringkasan dinyatakan valid karena mampu memuat informasi yang ada pada sumber dokumen. Persentase keakuratan hasil ringkasan akan dihitung dengan menginisialisasi similaritas inter dan intra. Similaritas inter merupakan rata-rata nilai similaritas hasil ringkasan dengan tiap dokumen dengan rumus

$$A = \text{cosine_sim}(\text{docA}, \text{summary}) \quad (3.2)$$

$$B = \text{cosine_sim}(\text{docB}, \text{summary}) \quad (3.3)$$

$$C = \text{cosine_sim}(\text{docC}, \text{summary}) \quad (3.4)$$

$$D = \text{cosine_sim}(\text{docD}, \text{summary}) \quad (3.5)$$

$$\text{inter_sim} = (A+B+C+D)/4 \quad (3.6)$$

dan intra similaritas adalah rata-rata nilai similaritas antar dokumen dengan rumus

$$E = \text{cosine_sim}(\text{docA}, \text{docB}) \quad (3.7)$$

$$F = \text{cosine_sim}(\text{docA}, \text{docC}) \quad (3.8)$$

$$G = \text{cosine_sim}(\text{docA}, \text{docD}) \quad (3.9)$$

$$H = \text{cosine_sim}(\text{docB}, \text{docC}) \quad (3.10)$$

$$I = \text{cosine_sim}(\text{docB}, \text{docD}) \quad (3.11)$$

$$J = \text{cosine_sim}(\text{docC}, \text{docD}) \quad (3.12)$$

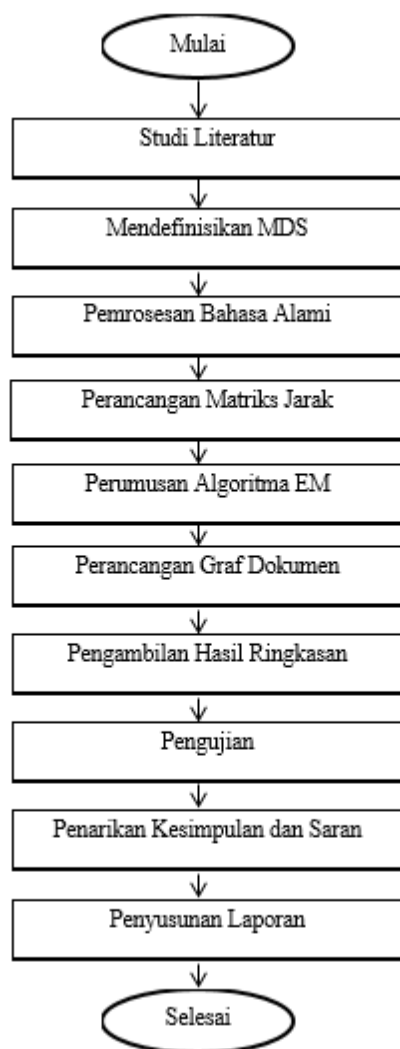
$$\text{intra_sim} = (E+F+G+H+I+J)/6 \quad (3.13)$$

3. Penarikan Kesimpulan dan Saran

Setelah dilakukan pengujian, maka dapat ditarik suatu kesimpulan dan saran sebagai masukan untuk pengembangan penelitian lebih lanjut.

4. Penyusunan Laporan Tugas Akhir

Setelah semua proses selesai dilakukan maka tahap terakhir adalah penyusunan laporan Tugas Akhir.



Gambar 3.1 Diagram Alir Metode Penelitian

BAB IV

PERANCANGAN IMPLEMENTASI

Bab ini menjelaskan rancangan implementasi yang menggambarkan proses rancang bangun secara terperinci dan digunakan sebagai acuan untuk implementasi.

4.1 Pemrosesan Bahasa Alami untuk Data Masukan

Tujuan pemrosesan bahasa alami adalah untuk memanipulasi teks berbasis sintaks (yang sering kali disebut dengan ‘word counting’) ke pemahaman yang benar dan memproses bahasa alami yang mempertimbangkan berbagai batasan semantik dan gramatikal dan juga konteks. Pemrosesan ini terdiri dari beberapa tahap yaitu [20],

1. *Split Sentence*

Pemisahan kalimat merupakan proses pemecahan dokumen menjadi kalimat-kalimat. Implementasi program untuk melakukan proses *Split Sentence* atau Pemisahan Kalimat dilakukan melakukan modul nltk sebagai berikut:

```
bowA = preprocess(docA).split(" ")
bowB = preprocess(docB).split(" ")
bowC = preprocess(docC).split(" ")
bowD = preprocess(docD).split(" ")
```

Contoh hasil proses pemisahan kalimat adalah sebagai berikut,
docA = " A chemical attack in a Syrian rebel-held town has sparked widespread international outrage and revulsion. At least 85 people, including many women and children, were killed in Douma on Saturday, according to a statement by rescue workers and medical staff. "

Menjadi:

docA = [‘A chemical attack in a Syrian rebel-held town has sparked widespread international outrage and revulsion’, ‘At least 85 people, including many women and children, were killed in Douma on

Saturday, according to a statement by rescue workers and medical staff ‘]

docB = “US President Donald Trump has lashed out at Syrian President Bashar al-Assad and his allies Russia and Iran over a suspected chemical attack, saying there will be a big price to pay. Medical sources say dozens of people were killed in an attack on the rebel-held town of Douma on Saturday.”

Menjadi:

docB = [‘US President Donald Trump has lashed out at Syrian President Bashar al-Assad and his allies Russia and Iran over a suspected chemical attack saying there will be a big price to pay’, ‘Medical sources say dozens of people were killed in an attack on the rebel-held town of Douma on Saturday’]

docC = “SHOCKING images of Syrian children left fighting for their lives have beamed around the world following an apparent gas attack on Saturday evening. More than 500 civilians — mostly women and children — are said to be injured following the attack in Douma, with many of those seen at medical centres having difficulty breathing. Local reports suggest up to 70 people may have been killed, including families found in their homes and shelters.”

Menjadi:

docC = [‘SHOCKING images of Syrian children left fighting for their lives have beamed around the world following an apparent gas attack on Saturday evening’, ‘More than 500 civilians — mostly women and children — are said to be injured following the attack in Douma, with many of those seen at medical centres having difficulty breathing’, ‘Local reports suggest up to 70 people may have been killed, including families found in their homes and shelters’]

docD = “At least 70 people have died in a suspected chemical attack in Douma, the last rebel-held town in Syria's Eastern Ghouta,

rescuers and medics say. Volunteer rescue force the White Helmets tweeted graphic images showing several bodies in basements. It said the deaths were likely to rise.”

Menjadi:

docD = [‘At least 70 people have died in a suspected chemical attack in Douma, the last rebel-held town in Syria's Eastern Ghouta, rescuers and medics say’, ‘Volunteer rescue force the White Helmets tweeted graphic images showing several bodies in basements’, ‘It said the deaths were likely to rise’]

2. Tokenisasi

Tokenisasi adalah proses untuk membagi teks yang dapat berupa kalimat, paragraf atau dokumen, menjadi token-token/bagian-bagian tertentu. Biasanya, yang menjadi acuan pemisah antar token adalah spasi dan tanda baca [20]. Implementasi program untuk melakukan proses Tokenisasi dilakukan menggunakan modul nltk sebagai berikut:

```
tokens = tokenizer.tokenize(sentence)
```

Contoh proses tahap tokenisasi atas data masukan adalah sebagai berikut.

docA = [‘A chemical attack in a Syrian rebel-held town has sparked widespread international outrage and revulsion’, ‘At least 85 people, including many women and children, were killed in Douma on Saturday, according to a statement by rescue workers and medical staff’]

Menjadi:

docA = ['A', 'chemical', 'attack', 'in', 'a', 'Syrian', 'rebel-held', 'town', 'has', 'sparked', 'widespread', 'international', 'outrage', 'and', 'revulsion.', 'At', 'least', '85', 'people,', 'including', 'many', 'women', 'and', 'children,', 'were', 'killed', 'in', 'Douma', 'on', 'Saturday,', 'according', 'to', 'a', 'statement', 'by', 'rescue', 'workers', 'and', 'medical', 'staff.']

docB = ['US President Donald Trump has lashed out at Syrian President Bashar al-Assad and his allies Russia and Iran over a suspected chemical attack saying there will be a big price to pay', 'Medical sources say dozens of people were killed in an attack on the rebel-held town of Douma on Saturday']

Menjadi:

docB = ['US', 'President', 'Donald', 'Trump', 'has', 'lashed', 'out', 'at', 'Syrian', 'President', 'Bashar', 'al-Assad', 'and', 'his', 'allies', 'Russia', 'and', 'Iran', 'over', 'a', 'suspected', 'chemical', 'attack,', 'saying', 'there', 'will', 'be', 'a', 'big', 'price', 'to', 'pay.', 'Medical', 'sources', 'say', 'dozens', 'of', 'people', 'were', 'killed', 'in', 'an', 'attack', 'on', 'the', 'rebel-held', 'town', 'of', 'Douma', 'on', 'Saturday.']

docC = ['SHOCKING images of Syrian children left fighting for their lives have beamed around the world following an apparent gas attack on Saturday evening', 'More than 500 civilians — mostly women and children — are said to be injured following the attack in Douma, with many of those seen at medical centres having difficulty breathing', 'Local reports suggest up to 70 people may have been killed, including families found in their homes and shelters']

Menjadi:

docC = ['SHOCKING', 'images', 'of', 'Syrian', 'children', 'left', 'fighting', 'for', 'their', 'lives', 'have', 'beamed', 'around', 'the', 'world', 'following', 'an', 'apparent', 'gas', 'attack', 'on', 'Saturday', 'evening.', 'More', 'than', '500', 'civilians', '—', 'mostly', 'women', 'and', 'children', '—', 'are', 'said', 'to', 'be', 'injured', 'following', 'the', 'attack', 'in', 'Douma,', 'with', 'many', 'of', 'those', 'seen', 'at', 'medical', 'centres', 'having', 'difficulty', 'breathing.', 'Local', 'reports', 'suggest', 'up', 'to', '70', 'people', 'may', 'have', 'been', 'killed,', 'including', 'families', 'found', 'in', 'their', 'homes', 'and', 'shelters.']

docD = ['At least 70 people have died in a suspected chemical attack in Douma, the last rebel-held town in Syria's Eastern Ghouta,

rescuers and medics say’, ‘Volunteer rescue force the White Helmets tweeted graphic images showing several bodies in basements’, ‘It said the deaths were likely to rise’]

Menjadi:

docD = ['At', 'least', '70', 'people', 'have', 'died', 'in', 'a', 'suspected', 'chemical', 'attack', 'in', 'Douma,', 'the', 'last', 'rebel-held', 'town', 'in', 'Syria's', 'Eastern', 'Ghouta,', 'rescuers', 'and', 'medics', 'say.', 'Volunteer', 'rescue', 'force', 'the', 'White', 'Helmets', 'tweeted', 'graphic', 'images', 'showing', 'several', 'bodies', 'in', 'basements.', 'It', 'said', 'the', 'deaths', 'were', 'likely', 'to', 'rise.']

3. *Stopword Removal*

Stopwords Removal adalah proses menghilangkan kata umum (common words) yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna [20]. Implementasi program untuk melakukan proses ini dilakukan menggunakan modul nltk sebagai berikut:

```
filtered_words = [w for w in tokens if not
w in stopwords.words('english')]
```

dan berikut adalah contoh proses tahap *Stopwords Removal* atas data masukan:

docA = ['A', 'chemical', 'attack', 'in', 'a', 'Syrian', 'rebel-held', 'town', 'has', 'sparked', 'widespread', 'international', 'outrage', 'and', 'revulsion.', 'At', 'least', '85', 'people,', 'including', 'many', 'women', 'and', 'children,', 'were', 'killed', 'in', 'Douma', 'on', 'Saturday,', 'according', 'to', 'a', 'statement', 'by', 'rescue', 'workers', 'and', 'medical', 'staff.']

Menjadi:

docA = ['chemical', 'attack', 'syrian', 'rebel', 'held', 'town', 'sparked', 'widespread', 'international', 'outrage', 'revulsion', 'least', '85', 'people', 'including', 'many', 'women', 'children', 'killed', 'douma', 'saturday', 'according', 'statement', 'rescue', 'workers', 'medical', 'staff']

docB = ['US', 'President', 'Donald', 'Trump', 'has', 'lashed', 'out', 'at', 'Syrian', 'President', 'Bashar', 'al-Assad', 'and', 'his', 'allies', 'Russia', 'and', 'Iran', 'over', 'a', 'suspected', 'chemical', 'attack', 'saying', 'there', 'will', 'be', 'a', 'big', 'price', 'to', 'pay.', 'Medical', 'sources', 'say', 'dozens', 'of', 'people', 'were', 'killed', 'in', 'an', 'attack', 'on', 'the', 'rebel-held', 'town', 'of', 'Douma', 'on', 'Saturday.']

Menjadi:

docB = ['us', 'president', 'donald', 'trump', 'lashed', 'syrian', 'president', 'bashar', 'al', 'assad', 'allies', 'russia', 'iran', 'suspected', 'chemical', 'attack', 'saying', 'big', 'price', 'pay', 'medical', 'sources', 'say', 'dozens', 'people', 'killed', 'attack', 'rebel', 'held', 'town', 'douma', 'saturday']

docC = ['SHOCKING', 'images', 'of', 'Syrian', 'children', 'left', 'fighting', 'for', 'their', 'lives', 'have', 'beamed', 'around', 'the', 'world', 'following', 'an', 'apparent', 'gas', 'attack', 'on', 'Saturday', 'evening.', 'More', 'than', '500', 'civilians', '—', 'mostly', 'women', 'and', 'children', '—', 'are', 'said', 'to', 'be', 'injured', 'following', 'the', 'attack', 'in', 'Douma,', 'with', 'many', 'of', 'those', 'seen', 'at', 'medical', 'centres', 'having', 'difficulty', 'breathing.', 'Local', 'reports', 'suggest', 'up', 'to', '70', 'people', 'may', 'have', 'been', 'killed,', 'including', 'families', 'found', 'in', 'their', 'homes', 'and', 'shelters.']

Menjadi:

docC = ['shocking', 'images', 'syrian', 'children', 'left', 'fighting', 'lives', 'beamed', 'around', 'world', 'following', 'apparent', 'gas', 'attack', 'saturday', 'evening', '500', 'civilians', 'mostly', 'women', 'children', 'said', 'injured', 'following', 'attack', 'douma', 'many', 'seen', 'medical', 'centres', 'difficulty', 'breathing', 'local', 'reports', 'suggest', '70', 'people', 'may', 'killed', 'including', 'families', 'found', 'homes', 'shelters']

docD = ['At', 'least', '70', 'people', 'have', 'died', 'in', 'a', 'suspected', 'chemical', 'attack', 'in', 'Douma,', 'the', 'last', 'rebel-held', 'town', 'in',

"Syria's", 'Eastern', 'Ghouta,', 'rescuers', 'and', 'medics', 'say.', 'Volunteer', 'rescue', 'force', 'the', 'White', 'Helmets', 'tweeted', 'graphic', 'images', 'showing', 'several', 'bodies', 'in', 'basements.', 'It', 'said', 'the', 'deaths', 'were', 'likely', 'to', 'rise.']

Menjadi:

docD = ['least', '70', 'people', 'died', 'suspected', 'chemical', 'attack', 'douma', 'last', 'rebel', 'held', 'town', 'syria', 'eastern', 'ghouta', 'rescuers', 'medics', 'say', 'volunteer', 'rescue', 'force', 'white', 'helmets', 'tweeted', 'graphic', 'images', 'showing', 'several', 'bodies', 'basements', 'said', 'deaths', 'likely', 'rise']

4.2 Perancangan Matriks Jarak

Pada tahap ini, proses perancangan model ruang vektor menggunakan nilai tf-idf dilakukan atas dokumen masukan pada subbab 4.1 menggunakan Persamaan (2.1) sebagai berikut

$$tf - idf_{t,d} = tf_{t,d} \cdot idf_t$$

dimana

$tf - idf_{t,d}$ = bobot TF/IDF kata_t pada dokumen_d

tf = frekuensi kata_t pada dokumen_d

idf_t = nilai idf dari kata_t

dengan

$$idf_{t,d} = \log \frac{N}{df_t}$$

Implementasi program untuk melakukan proses penghitungan nilai TF-IDF adalah sebagai berikut:

```
def computeTFIDF(tfBow, idfs):
    tfidf={}
    for word, val in tfBow.items():
        tfidf[word]=val*idfs[word]
    return tfidf

tfidfBowA = computeTFIDF(tfBowA, idfs)
tfidfBowB = computeTFIDF(tfBowB, idfs)
```

```
tfidfBowC = computeTFIDF(tfBowC, idfs)
tfidfBowD = computeTFIDF(tfBowD, idfs)
```

dan pada tabel 4.1 di bawah ini diperoleh hasil pencarian nilai TF-IDF dari tiap kata yang ada pada seluruh sumber dokumen.

Tabel 4.1 Contoh Nilai TF-IDF Kata

Kata	docA	docB	docC	docD
500	0.000000	0.000000	0.054234	0.000000
70	0.000000	0.000000	0.038481	0.049798
85	0.088381	0.000000	0.000000	0.000000
according	0.088381	0.000000	0.000000	0.000000
al	0.000000	0.074572	0.000000	0.000000
allies	0.000000	0.074572	0.000000	0.000000
apparent	0.000000	0.000000	0.054234	0.000000
around	0.000000	0.000000	0.054234	0.000000
assad	0.000000	0.074572	0.000000	0.000000
attack	0.037037	0.062500	0.045455	0.029412

Setelah mendapatkan nilai tf-idf untuk setiap kata, dihitung jarak Euclidean pada Persamaan (2.3). Jarak yang dihitung adalah jarak dari tiap kata pada dokumen terhadap titik tengah yang telah ditentukan pada tahap *Guess Model Parameter* dalam proses perumusan algoritma EM. Matriks jarak merupakan input untuk tahap clustering. Implementasi program untuk melakukan proses penghitungan matriks jarak adalah sebagai berikut:

```
def dist(x,y):
    return numpy.sqrt(numpy.sum((x-y)**2))
dist(a,b)
dist(a,c)
```

Didapat nilai jarak masing-masing kata terhadap cluster yang tertera pada Tabel 4.2 sebagai berikut:

Tabel 4.2 Contoh Nilai Jarak Kata terhadap Titik Tengah Tiap Cluster

Kata	Jarak terhadap Titik a	Jarak terhadap Titik b
500	0.145766	0.287995358914
70	0.169021383751	0.253064079168
85	0.218657726049	0.229038863866
according	0.218657726049	0.229038863866
al	0.213450189	0.292508090801
allies	0.213450189	0.292508090801
apparent	0.145766	0.287995358914
around	0.145766	0.287995358914
assad	0.213450189	0.292508090801
attack	0.173283617627	0.248253121104

4.3 Perumusan Algoritma Ekspektasi Maksimisasi

Dalam setiap iterasi pada Algoritma EM ini terdapat 2 tahap, yaitu tahap Ekspektasi atau tahap E (E step) dan tahap Maksimisasi atau tahap M (M step).

Langkah-langkah algoritma Ekspektasi Maksimisasi untuk clustering kata adalah sebagai berikut:

1. *Guess Model Parameter*

Menginisialisasi secara acak titik tengah untuk tiap cluster. Sebagai contoh, untuk meringkas 4 dokumen yang ada pada subbab 4.1 ditetapkan 2 titik tengah a dan b karena akan dilakukan clustering dengan 2 cluster.

$$a = (0.000000, 0.000000, 0.200000, 0.000000)$$

$$b = (0.200000, 0.000000, 0.000000, 0.200000)$$

Implementasi program untuk melakukan proses *Guess Model Parameter* adalah sebagai berikut:

```
bbr=numpy.array((0.0,0.0,0.2,0.0))
cbr=numpy.array((0.2,0.0,0.0,0.2))
```

2. Tahap Ekspektasi

Menghitung bobot keanggotaan tiap kata pada dokumen terhadap cluster 1 dengan rumus pada Persamaan (2.4) sebagai berikut

$$w_{o,c_1} = \frac{(dist(o, c_2))^2}{(dist(o, c_1))^2 + (dist(o, c_1))^2}$$

bobot keanggotaan tiap kata pada dokumen terhadap cluster 2 dengan rumus pada Persamaan (2.5) sebagai berikut

$$w_{o,c_2} = \frac{(dist(o, c_1))^2}{(dist(o, c_1))^2 + (dist(o, c_1))^2}$$

didapat bobot keanggotaan tiap kata pada dokumen terhadap cluster 1 dan cluster 2. Kata dengan bobot keanggotaan cluster 1 yang lebih besar termasuk ke dalam cluster 1. Sedangkan kata dengan bobot keanggotaan cluster 2 yang lebih besar termasuk ke dalam cluster 2. Implementasi program untuk melakukan penghitungan bobot keanggotaan pada tahap ekspektasi adalah sebagai berikut:

```
for number in range (0,w.size):
    c1 = eu[number]
    c2 = ac[number]
    c = c1+c2
    wsatu = ac[number]/c
    wdua = eu[number]/c

    if wsatu>wdua:
        cluster1 =
numpy.append(cluster1,words[number])
    else:
        cluster2 =
numpy.append(cluster2,words[number])
```

Hasil penghitungan bobot keanggotaan pada Tabel 2.2 dari dokumen masukan pada subbab 4.1 pada iterasi terakhir tertera pada Tabel 4.3 sebagai berikut:

Tabel 4.3 Contoh Bobot Keanggotaan Kata terhadap Cluster pada Iterasi Terakhir

Bobot Keanggotaan Cluster 1	Bobot Keanggotaan Cluster 2
0.50000481	0.49999519
0.49999845	0.50000155
0.50000018	0.49999982
0.50001034	0.49998966
0.50000018	0.49999982
0.50001034	0.49998966
0.49999649	0.50000351
0.50000018	0.49999982
0.49999842	0.50000158
0.50001034	0.49998966

Dari Tabel 4.2 di atas didapat hasil clustering kata pada iterasi terakhir sebagai berikut

Cluster 1	Cluster 2
['many' 'international' 'apparent' '85' 'injured' 'statement' 'shocking' 'douma' 'outrage' 'women' 'found' '70' 'children' 'people' 'medical' 'syrian' 'left' 'local' 'workers' 'attack' 'mostly' 'killed' 'sparked' 'following' 'fighting' 'suggest' 'evening' 'reports' 'difficulty' 'revulsion' 'images' 'gas' 'world' 'according' 'including' 'lives' 'seen' 'around' 'widespread' 'beamed' 'may' 'civilians' '500' 'said' 'shelters']	['least' 'eastern' 'assad' 'suspected' 'helmets' 'rescue' 'several' 'big' 'died' 'price' 'volunteer' 'bodies' 'iran' 'rise' 'deaths' 'rebel' 'lashed' 'bashar' 'trump' 'held' 'syria' 'rescuers' 'allies' 'white' 'graphic' 'chemical' 'likely' 'force' 'al' 'showing' 'president' 'tweeted' 'saying' 'russia' 'pay' 'say' 'town' 'basements' 'donald' 'medics' 'us' 'last' 'sources' 'ghouta' 'dozens']

'families' 'staff' 'saturday' 'breathing' 'homes' 'centres']	
---	--

3. Tahap Maksimisasi

Pada tahap ini, dilakukan pencarian titik tengah baru yang lebih sesuai dengan persebaran kata pada tiap kluster pada Persamaan (2.6) dengan rumus sebagai berikut

$$c_j = \frac{\sum_{\text{each point } o} w_{o,c_j}^2 o}{\sum_{\text{each point } o} w_{o,c_j}^2}$$

Implementasi program untuk mencari titik tengah baru masing-masing cluster pada tahap maksimisasi adalah sebagai berikut:

```
for doc in range(0,4):
    cl = numpy.array([],dtype=float)
    for s in wordset:
        p = k[s][doc]
        cl = numpy.append(cl,p)
    p = (tabel1**2)*cl
    q = (tabel2**2)*cl
    bnew = p.sum()/(tabel1.sum())
    bbr = numpy.append(bbr,bnew)
    cnew = q.sum()/(tabel1.sum())
    cbr = numpy.append(cbr,cnew)
```

dan sebagai contoh, didapat hasil nilai titik tengah baru pada iterasi terakhir atas dokumen masukan pada 4.1 sebagai berikut

$$a_{br} = (0.00912872 \ 0.00986988 \ 0.01059597 \ 0.01037495)$$

$$b_{br} = (0.00912864 \ 0.00986996 \ 0.01059526 \ 0.01037516)$$

4. Ulangi langkah 2 dan 3 sampai konvergen

Langkah ekspektasi dan maksimisasi terus dilakukan sampai

$$\sum_{i=0}^n |a_i - a_{bri}| \leq 0.000001$$

$$\sum_{i=0}^n |b_i - b_{bri}| \leq 0.000001$$

dimana n adalah banyaknya dokumen.

Implementasi program untuk terus melakukan pengulangan sampai nilai titik tengah tiap cluster tidak berubah adalah sebagai berikut:

```
while (abs(b-bbr)).sum()>0.000001 and
(abs(c-cbr)).sum()>0.000001:
```

Pada tiap iterasi, bobot keanggotaan dihitung kembali sesuai dengan titik tengah baru yang di dapat. Sebagai contoh, pada kasus 4.1 pengulangan tahap maksimisasi untuk mendapatkan nilai tengah terbaik adalah sebagai berikut:

Tabel 4.4 Contoh Pencarian Titik Tengah Terbaik

Iterasi	Titik Tengah Cluster 1 (a)	Titik Tengah Cluster 2 (b)
Ke-0	[0.000000 0.000000 0.200000 0.000000]	[0.200000 0.000000 0.000000 0.200000]
Ke-1	[0.00901313 0.0110243 0.01554307 0.010178]	[0.00683226 0.00627507 0.00445955 0.00775275]
Ke-2	[0.00920266 0.01001711 0.0119671 0.01019691]	[0.00837984 0.00898833 0.00855992 0.00977308]
Ke-3	[0.00916677 0.0098715 0.01103033 0.01027724]	[0.00889152 0.00965215 0.00994178 0.01024526]
Ke-4	[0.00914202 0.00985883 0.01074132 0.0103306]	[0.00905567 0.00981638 0.01038199 0.01035157]
Ke-5	[0.00913263 0.00986311 0.01064606 0.01035656]	[0.00910635 0.00985684 0.01052401 0.01037265]
Ke-6	[0.00912973 0.00986677 0.01061347 0.01036773]	[0.00912182 0.00986678 0.01057104 0.01037577]
Ke-7	[0.00912894 0.00986862 0.01060205 0.01037225]	[0.00912656 0.00986921 0.01058703 0.01037574]

Ke-8	[0.00912876 0.00986942 0.01059798 0.01037402]	[0.00912803 0.00986979 0.01059258 0.01037544]
Ke-9	[0.00912872 0.00986975 0.01059651 0.01037469]	[0.00912849 0.00986993 0.01059455 0.01037525]
Ke-10	[0.00912872 0.00986988 0.01059597 0.01037495]	[0.00912864 0.00986996 0.01059526 0.01037516]

Setelah mendapat bobot keanggotaan pada iterasi pertama, dapat ditemukan titik tengah baru untuk setiap iterasi (Tabel 4.4). Titik tengah baru tersebut digunakan untuk menghitung kembali bobot keanggotaan. Berikut Tabel 4.5 yang merupakan hasil cluster kata pada tiap iterasi yang bersesuaian dengan Tabel 2.2.

Tabel 4.5 Contoh Hasil Clustering Kata Pada tiap Iterasi

Iterasi	Cluster 1	Cluster 2
Ke-0	['many' 'international' 'apparent' '85' 'injured' 'eastern' 'statement' 'assad' 'shocking' 'douma' 'outrage' 'women' 'suspected' 'helmets' 'found' '70' 'children' 'several' 'people' 'big' 'died' 'price' 'medical' 'volunteer' 'syrian' 'bodies' 'iran' 'left' 'rise' 'deaths' 'rebel' 'local' 'lashed' 'workers' 'bashar' 'attack' 'trump' 'mostly' 'killed' 'sparked' 'following' 'held' 'fighting' 'syria' 'rescuers' 'suggest' 'evening' 'reports' 'difficulty' 'revulsion' 'allies' 'images' 'white' 'graphic' 'chemical' 'gas' 'likely' 'world' 'force' 'according' 'al' 'including' 'lives' 'seen' 'showing' 'president' 'tweeted' 'saying' 'around' 'russia' 'widespread' 'pay' 'say' 'beamed' 'may' 'civilians']	['least' 'rescue']

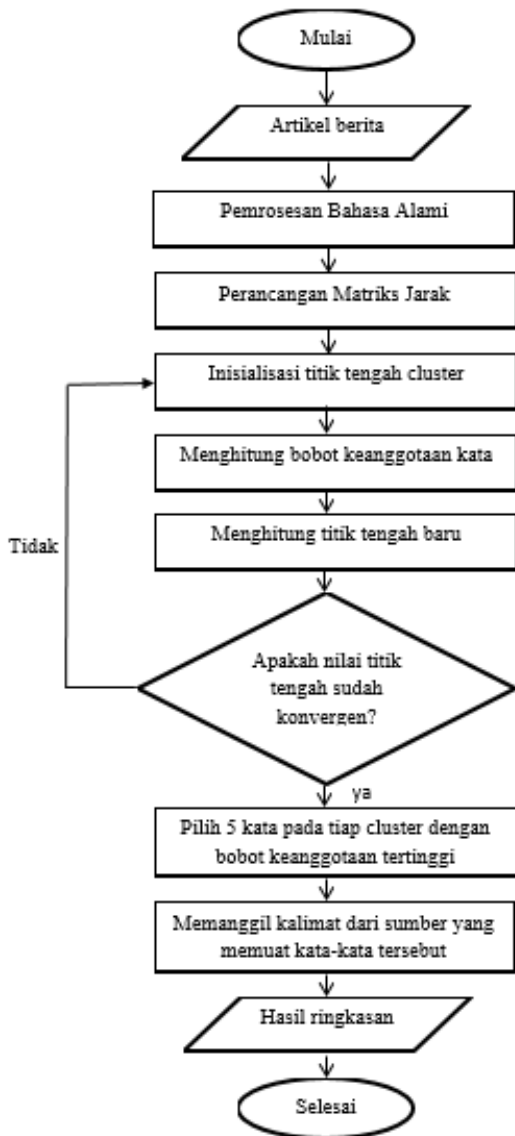
	'500' 'said' 'town' 'shelters' 'families' 'basements' 'staff' 'donald' 'medics' 'us' 'last' 'saturday' 'sources' 'ghouta' 'dozens' 'breathing' 'homes' 'centres']	
Ke-1	['many' 'least' 'international' 'apparent' '85' 'injured' 'statement' 'assad' 'shocking' 'douma' 'outrage' 'women' 'suspected' 'found' 'rescue' '70' 'children' 'people' 'big' 'price' 'medical' 'syrian' 'iran' 'left' 'rebel' 'local' 'lashed' 'workers' 'bashar' 'attack' 'trump' 'mostly' 'killed' 'sparked' 'following' 'held' 'fighting' 'suggest' 'evening' 'reports' 'difficulty' 'revulsion' 'allies' 'images' 'chemical' 'gas' 'world' 'according' 'al' 'including' 'lives' 'seen' 'president' 'saying' 'around' 'russia' 'widespread' 'pay' 'say' 'beamed' 'may' 'civilians' '500' 'said' 'town' 'shelters' 'families' 'staff' 'donald' 'us' 'saturday' 'sources' 'dozens' 'breathing' 'homes' 'centres']	['eastern' 'helmets' 'several' 'died' 'volunteer' 'bodies' 'rise' 'deaths' 'syria' 'rescuers' 'white' 'graphic' 'likely' 'force' 'showing' 'tweeted' 'basements' 'medics' 'last' 'ghouta']
Ke-2	['many' 'least' 'international' 'apparent' '85' 'injured' 'statement' 'assad' 'shocking' 'douma' 'outrage' 'women' 'suspected' 'found' 'rescue' '70' 'children' 'people' 'big' 'price' 'medical' 'syrian' 'iran' 'left' 'rebel' 'local' 'lashed' 'workers' 'bashar' 'attack' 'trump' 'mostly'	['eastern' 'helmets' 'several' 'died' 'volunteer' 'bodies' 'rise' 'deaths' 'syria' 'rescuers' 'white' 'graphic' 'likely' 'force' 'showing' 'tweeted' 'basements' 'medics' 'last' 'ghouta']

	'killed' 'sparked' 'following' 'held' 'fighting' 'suggest' 'evening' 'reports' 'difficulty' 'revulsion' 'allies' 'images' 'chemical' 'gas' 'world' 'according' 'al' 'including' 'lives' 'seen' 'president' 'saying' 'around' 'russia' 'widespread' 'pay' 'say' 'beamed' 'may' 'civilians' '500' 'said' 'town' 'shelters' 'families' 'staff' 'donald' 'us' 'saturday' 'sources' 'dozens' 'breathing' 'homes' 'centres']	
Ke-3	['many' 'least' 'international' 'apparent' '85' 'injured' 'statement' 'shocking' 'douma' 'outrage' 'women' 'found' 'rescue' '70' 'children' 'people' 'medical' 'syrian' 'left' 'rebel' 'local' 'workers' 'attack' 'mostly' 'killed' 'sparked' 'following' 'held' 'fighting' 'suggest' 'evening' 'reports' 'difficulty' 'revulsion' 'images' 'chemical' 'gas' 'world' 'according' 'including' 'lives' 'seen' 'president' 'around' 'widespread' 'beamed' 'may' 'civilians' '500' 'said' 'town' 'shelters' 'families' 'staff' 'saturday' 'breathing' 'homes' 'centres']	['eastern' 'assad' 'suspected' 'helmets' 'several' 'big' 'died' 'price' 'volunteer' 'bodies' 'iran' 'rise' 'deaths' 'lashed' 'bashar' 'trump' 'syria' 'rescuers' 'allies' 'white' 'graphic' 'likely' 'force' 'al' 'showing' 'tweeted' 'saying' 'russia' 'pay' 'say' 'basements' 'donald' 'medics' 'us' 'last' 'sources' 'ghouta' 'dozens']
Ke-4	['many' 'international' 'apparent' '85' 'injured' 'statement' 'shocking' 'douma' 'outrage' 'women' 'found' '70' 'children' 'people' 'medical' 'syrian' 'left' 'rebel' 'local' 'workers' 'attack' 'mostly' 'killed'	['least' 'eastern' 'assad' 'suspected' 'helmets' 'rescue' 'several' 'big' 'died' 'price' 'volunteer' 'bodies' 'iran' 'rise' 'deaths' 'lashed'

	'sparked' 'following' 'held' 'fighting' 'suggest' 'evening' 'reports' 'difficulty' 'revulsion' 'images' 'chemical' 'gas' 'world' 'according' 'including' 'lives' 'seen' 'president' 'around' 'widespread' 'beamed' 'may' 'civilians' '500' 'said' 'town' 'shelters' 'families' 'staff' 'saturday' 'breathing' 'homes' 'centres']	'bashar' 'trump' 'syria' 'rescuers' 'allies' 'white' 'graphic' 'likely' 'force' 'al' 'showing' 'tweeted' 'saying' 'russia' 'pay' 'say' 'basements' 'donald' 'medics' 'us' 'last' 'sources' 'ghouta' 'dozens']
Ke-5	['many' 'international' 'apparent' '85' 'injured' 'statement' 'shocking' 'douma' 'outrage' 'women' 'found' '70' 'children' 'people' 'medical' 'syrian' 'left' 'local' 'workers' 'attack' 'mostly' 'killed' 'sparked' 'following' 'fighting' 'suggest' 'evening' 'reports' 'difficulty' 'revulsion' 'images' 'gas' 'world' 'according' 'including' 'lives' 'seen' 'around' 'widespread' 'beamed' 'may' 'civilians' '500' 'said' 'shelters' 'families' 'staff' 'saturday' 'breathing' 'homes' 'centres']	['least' 'eastern' 'assad' 'suspected' 'helmets' 'rescue' 'several' 'big' 'died' 'price' 'volunteer' 'bodies' 'iran' 'rise' 'deaths' 'rebel' 'lashed' 'bashar' 'trump' 'held' 'syria' 'rescuers' 'allies' 'white' 'graphic' 'chemical' 'likely' 'force' 'al' 'showing' 'president' 'tweeted' 'saying' 'russia' 'pay' 'say' 'town' 'basements' 'donald' 'medics' 'us' 'last' 'sources' 'ghouta' 'dozens']
Ke-6	['many' 'international' 'apparent' '85' 'injured' 'statement' 'shocking' 'douma' 'outrage' 'women' 'found' '70' 'children' 'people' 'medical' 'syrian' 'left' 'local' 'workers' 'attack' 'mostly' 'killed' 'sparked' 'following' 'fighting' 'suggest' 'evening' 'reports' 'difficulty' 'revulsion' 'images' 'gas' 'world' 'according' 'including' 'lives' 'seen'	['least' 'eastern' 'assad' 'suspected' 'helmets' 'rescue' 'several' 'big' 'died' 'price' 'volunteer' 'bodies' 'iran' 'rise' 'deaths' 'rebel' 'lashed' 'bashar' 'trump' 'held' 'syria' 'rescuers' 'allies' 'white'

	'around' 'widespread' 'beamed' 'may' 'civilians' '500' 'said' 'shelters' 'families' 'staff' 'saturday' 'breathing' 'homes' 'centres']	'graphic' 'chemical' 'likely' 'force' 'al' 'showing' 'president' 'tweeted' 'saying' 'russia' 'pay' 'say' 'town' 'basements' 'donald' 'medics' 'us' 'last' 'sources' 'ghouta' 'dozens']
Ke-7	['many' 'international' 'apparent' '85' 'injured' 'statement' 'shocking' 'douma' 'outrage' 'women' 'found' '70' 'children' 'people' 'medical' 'syrian' 'left' 'local' 'workers' 'attack' 'mostly' 'killed' 'sparked' 'following' 'fighting' 'suggest' 'evening' 'reports' 'difficulty' 'revulsion' 'images' 'gas' 'world' 'according' 'including' 'lives' 'seen' 'around' 'widespread' 'beamed' 'may' 'civilians' '500' 'said' 'shelters' 'families' 'staff' 'saturday' 'breathing' 'homes' 'centres']	['least' 'eastern' 'assad' 'suspected' 'helmets' 'rescue' 'several' 'big' 'died' 'price' 'volunteer' 'bodies' 'iran' 'rise' 'deaths' 'rebel' 'lashed' 'bashar' 'trump' 'held' 'syria' 'rescuers' 'allies' 'white' 'graphic' 'chemical' 'likely' 'force' 'al' 'showing' 'president' 'tweeted' 'saying' 'russia' 'pay' 'say' 'town' 'basements' 'donald' 'medics' 'us' 'last' 'sources' 'ghouta' 'dozens']
Ke-8	['many' 'international' 'apparent' '85' 'injured' 'statement' 'shocking' 'douma' 'outrage' 'women' 'found' '70' 'children' 'people' 'medical' 'syrian' 'left' 'local' 'workers' 'attack' 'mostly' 'killed' 'sparked' 'following' 'fighting' 'suggest' 'evening' 'reports' 'difficulty' 'revulsion' 'images' 'gas' 'world' 'according' 'including' 'lives' 'seen' 'around' 'widespread' 'beamed' 'may' 'civilians' '500' 'said' 'shelters'	['least' 'eastern' 'assad' 'suspected' 'helmets' 'rescue' 'several' 'big' 'died' 'price' 'volunteer' 'bodies' 'iran' 'rise' 'deaths' 'rebel' 'lashed' 'bashar' 'trump' 'held' 'syria' 'rescuers' 'allies' 'white' 'graphic' 'chemical' 'likely' 'force' 'al' 'showing' 'president' 'tweeted'

	'families' 'staff' 'saturday' 'breathing' 'homes' 'centres']	'saying' 'russia' 'pay' 'say' 'town' 'basements' 'donald' 'medics' 'us' 'last' 'sources' 'ghouta' 'dozens']
Ke-9	['many' 'international' 'apparent' '85' 'injured' 'statement' 'shocking' 'douma' 'outrage' 'women' 'found' '70' 'children' 'people' 'medical' 'syrian' 'left' 'local' 'workers' 'attack' 'mostly' 'killed' 'sparked' 'following' 'fighting' 'suggest' 'evening' 'reports' 'difficulty' 'revulsion' 'images' 'gas' 'world' 'according' 'including' 'lives' 'seen' 'around' 'widespread' 'beamed' 'may' 'civilians' '500' 'said' 'shelters' 'families' 'staff' 'saturday' 'breathing' 'homes' 'centres']	['least' 'eastern' 'assad' 'suspected' 'helmets' 'rescue' 'several' 'big' 'died' 'price' 'volunteer' 'bodies' 'iran' 'rise' 'deaths' 'rebel' 'lashed' 'bashar' 'trump' 'held' 'syria' 'rescuers' 'allies' 'white' 'graphic' 'chemical' 'likely' 'force' 'al' 'showing' 'president' 'tweeted' 'saying' 'russia' 'pay' 'say' 'town' 'basements' 'donald' 'medics' 'us' 'last' 'sources' 'ghouta' 'dozens']



Gambar 4.1 Diagram Alir Algoritma EM untuk Peringkasan Multi-Dokumen

4.4 Perancangan Graf Dokumen

Pada proses ini, 5 kata teratas dengan bobot keanggotaan tertinggi pada masing-masing cluster diambil untuk menyusun hasil ringkasan. Kalimat yang memuat kata dengan bobot keanggotaan tertinggi lah yang menjadi hasil ringkasan. Implementasi program untuk memanggil hasil ringkasan adalah sebagai berikut:.

```

for i in range(0,cl.size):
    for j in range(0,sa.size):
        if cluster2[y][i] in sentA[j] or
cluster1[x][i] in sentA[j]:

            sentfromA =
numpy.append(sentfromA,sentA[j])

for i in range(0,cl.size):
    for j in range(0,sb.size):
        if cluster2[y][i] in sentB[j] or
cluster1[x][i] in sentB[j]:
            sentfromB =
numpy.append(sentfromB,sentB[j])

for i in range(0,cl.size):
    for j in range(0,sc.size):
        if cluster2[y][i] in sentC[j] or
cluster1[x][i] in sentC[j]:
            sentfromC =
numpy.append(sentfromC,sentC[j])

for i in range(0,cl.size):
    for j in range(0,sd.size):
        if cluster2[y][i] in sentD[j] or
cluster1[x][i] in sentD[j]:
            sentfromD =
numpy.append(sentfromD,sentD[j])

summary =
set(sentfromA).union(set(sentfromB)).union(s
et(sentfromC)).union(set(sentfromD))

```

Pada dokumen masukkan di subbab 4.1, 5 kata dengan bobot keanggotaan tertinggi tiap cluster adalah

Cluster 1: ['world' 'workers' 'women' 'widespread' 'syrian']

Cluster 2: ['white' 'volunteer' 'us' 'tweeted' 'trump']

Kalimat pada tiap dokumen yang mengandung 10 kata tersebut dipanggil untuk menjadi hasil ringkasan

Hasil Ringkasan:

{'US President Donald Trump has lashed out at Syrian President Bashar al-Assad and his allies Russia and Iran over a suspected chemical attack, saying there will be a big price to pay', ' At least 85 people, including many women and children, were killed in Douma on Saturday, according to a statement by rescue workers and medical staff', 'A chemical attack in a Syrian rebel-held town has sparked widespread international outrage and revulsion', ' Volunteer rescue force the White Helmets tweeted graphic images showing several bodies in basements', 'SHOCKING images of Syrian children left fighting for their lives have beamed around the world following an apparent gas attack on Saturday evening'}

BAB V

IMPLEMENTASI DAN PENGUJIAN

Pada bab ini dijelaskan tentang implementasi dalam bahasa pemrograman Python dan hasil uji coba program dalam meringkas beberapa dokumen yang memiliki kemiripan topik.

5.1 Implementasi Perangkat Lunak

Program untuk menyelesaikan proses peringkasan diimplementasikan dalam bahasa pemrograman Python dengan menggunakan perangkat lunak Anaconda dan aplikasi Jupyter Notebook 5.0.0.

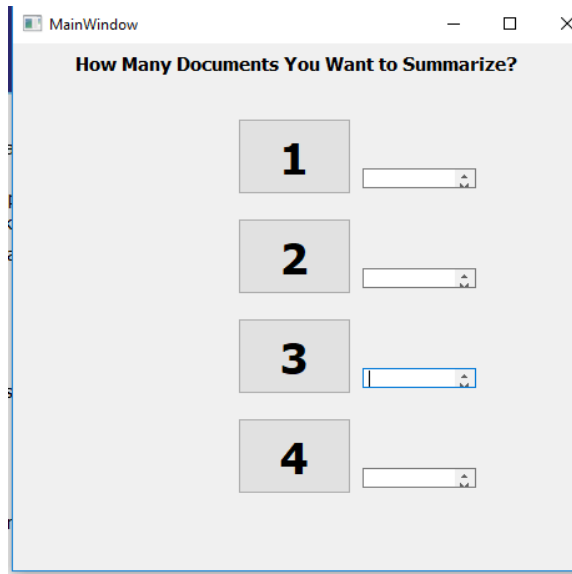
5.1.1 Tampilan Aplikasi

Ketika program dijalankan, muncul halaman utama dengan tombol *Summarize with String* dan *Summarize with Links* untuk memilih tipe data yang akan diinputkan beserta keterangan pemakaiannya. Tombol *Summarize with String* dipilih ketika pengguna ingin memasukkan artikel berita dengan melakukan *copy paste* artikel. Sedangkan tombol *Summarize with Links* dipilih ketika pengguna ingin menggunakan *link* artikel berita sebagai data masukan. Kemudian ketika pengguna telah menentukan satu dari dua tombol tersebut, *window* baru yang menanyakan jumlah dokumen yang ingin diproses akan terbuka. *Window* ini dilengkapi dengan 3 tombol yaitu tombol 2, 3, dan 4. Tombol 2, 3, dan 4, akan membuka *window* baru untuk memulai proses peringkasan. *Window* ini terdiri dari beberapa tombol dan isian untuk memudahkan proses peringkasan. Jumlah isian disesuaikan dengan pilihan pengguna pada *window* sebelumnya. Jika pengguna memilih tombol 2, maka isian pada *window* ini berjumlah 2, dan seterusnya. Isian ditandai dengan teks *place holder* yang menerangkan bagaimana pengguna mengisi isian tersebut sebagai data inputan. Tombol yang tersedia adalah tombol *Summarize now*. Tombol ini memeritahkan program untuk memproses ringkasan

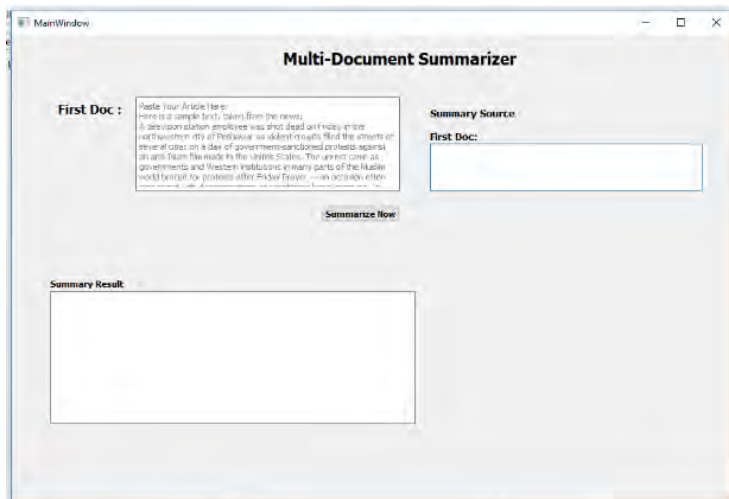
dari data masukan. Hasil ringkasan dan sumber ringkasan akan muncul pada *window* yang sama. Contoh tampilan program dapat dilihat pada gambar berikut.



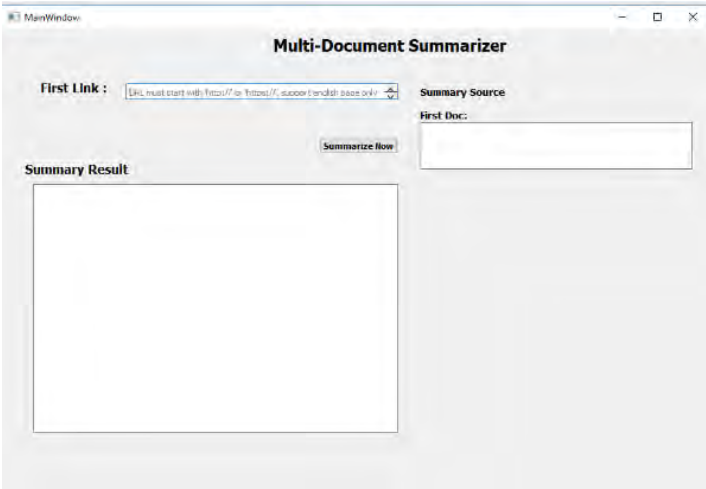
Gambar 5.1 Halaman Utama Program



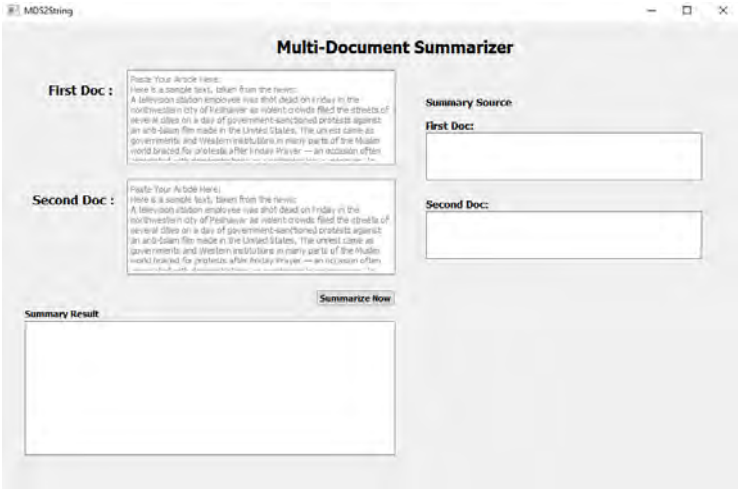
Gambar 5.2 Jendela *HowManyDocs*



Gambar 5.3 Jendela Peringkasan 1 Dokumen Menggunakan *String*



Gambar 5.4 Jendela Peringkasan 1 Dokumen Menggunakan *Link*



Gambar 5.5 Jendela Peringkasan 2 Dokumen Menggunakan *String*

The screenshot shows a window titled "Multi-Document Summarizer" with a tab labeled "MDS2Link". The interface includes the following elements:

- First Link :** A text input field containing a URL: `URL: first doc with https:// or http://? support length base: 1024`.
- Second Link :** A text input field containing a URL: `URL: second doc with https:// or http://? support length base: 1024`.
- Summary Source:** A section on the right with two empty text boxes labeled **First Doc:** and **Second Doc:**.
- Summary Result:** A large empty text box on the left.
- Buttons:** A "Summarize Now" button is located below the "Second Link" field.

Gambar 5.6 Jendela Peringkasan 2 Dokumen Menggunakan *Link*

The screenshot shows a window titled "Multi-Document Summarizer" with a tab labeled "MDS2String". The interface includes the following elements:

- First Doc :** A text input field containing sample text: `Paste Your Article Here: Here is a sample text, taken from the news: A television station employee was shot dead on Friday in the northwestern city of Baghdad as violent protests flared the streets of several cities on a day of government-sanctioned protests against an anti-corruption campaign.`
- Second Doc :** A text input field containing sample text: `Paste Your Article Here: Here is a sample text, taken from the news: A television station employee was shot dead on Friday in the northwestern city of Baghdad as violent protests flared the streets of several cities on a day of government-sanctioned protests against an anti-corruption campaign.`
- Third Doc :** A text input field containing sample text: `Paste Your Article Here: Here is a sample text, taken from the news: A television station employee was shot dead on Friday in the northwestern city of Baghdad as violent protests flared the streets of several cities on a day of government-sanctioned protests against an anti-corruption campaign.`
- Summary Source:** A section on the right with three empty text boxes labeled **First Doc:**, **Second Doc:**, and **Third Doc:**.
- Summary Result:** A large empty text box on the left.
- Buttons:** A "Summarize Now" button is located below the "Third Doc" field.

Gambar 5.7 Jendela Peringkasan 3 Dokumen Menggunakan *String*

The screenshot shows a web application titled "Multi-Document Summarizer". On the left, there are three input fields labeled "First Link:", "Second Link:", and "Third Link:". Each field contains a placeholder URL: "http://www.example.com/". To the right of these fields is a "Summary Source" section with three empty text boxes labeled "First Doc:", "Second Doc:", and "Third Doc:". Below the links is a "Summary Result" section with a large empty box. A "Summarize Now" button is located between the links and the summary result box.

Gambar 5.8 Jendela Peringkasan 3 Dokumen Menggunakan *Link*

The screenshot shows the same "Multi-Document Summarizer" application, but now the "First Doc:", "Second Doc:", "Third Doc:", and "Fourth Doc:" fields are populated with sample text. The text is repeated for each of the four document fields. The "Summary Result" box remains empty. The "Summarize Now" button is still present.

Gambar 5.9 Jendela Peringkasan 4 Dokumen Menggunakan *String*

Multi-Document Summarizer

First Link :

Second Link :

Third Link :

Fourth Link :

Summarize Now

Summary Result

Summary Source

First Doc:

Second Doc:

Third Doc:

Fourth Doc:

Gambar 5.10 **Jendela Peringkasan 4 Dokumen**
Menggunakan *Link*

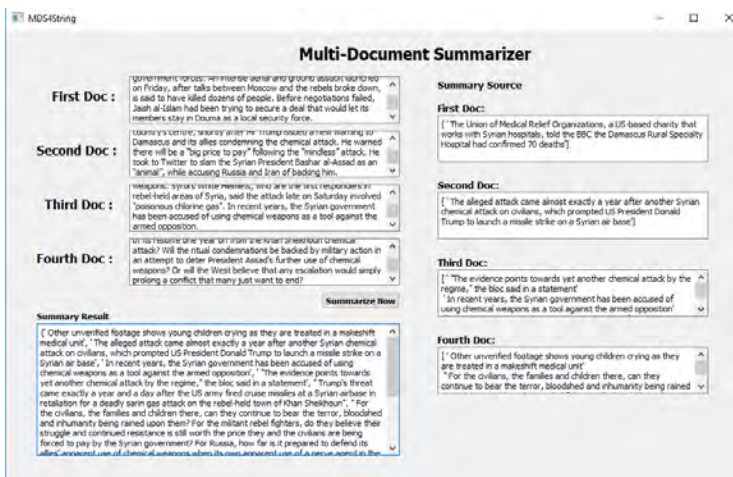
5.1.2 Pengujian Perangkat Lunak

Hasil dari peringkasan multi-dokumen dengan 4 dokumen masukan adalah sebagai berikut

Hasil Ringkasan:

{ ' Other unverified footage shows young children crying as they are treated in a makeshift medical unit', ' The alleged attack came almost exactly a year after another Syrian chemical attack on civilians, which prompted US President Donald Trump to launch a missile strike on a Syrian air base', ' In recent years, the Syrian government has been accused of using chemical weapons as a tool against the armed opposition', ' "The evidence points towards yet another chemical attack by the regime," the bloc said in a statement', " Trump's threat came exactly a year and a day after the US army fired cruise missiles at a Syrian airbase in retaliation for a deadly sarin gas attack on the rebel-held town of Khan Sheikhoun", " For the civilians, the families and children there, can they continue to bear the terror, bloodshed and inhumanity being rained upon them? For the militant rebel

fighters, do they believe their struggle and continued resistance is still worth the price they and the civilians are being forced to pay by the Syrian government? For Russia, how far is it prepared to defend its allies' apparent use of chemical weapons when its own apparent use of a nerve agent in the UK is subject to so much global condemnation? And for the international community, how will it respond to what is clearly a test of its resolve one year on from the Khan Sheikhoun chemical attack? Will the ritual condemnations be backed by military action in an attempt to deter President Assad's further use of chemical weapons? Or will the West believe that any escalation would simply prolong a conflict that many just want to end?", "" "Many dead, including women and children, in mindless CHEMICAL attack in Syria," Trump wrote on Twitter, lashing out at Syrian President Bashar al-Assad and his ally, Russia's Vladimir Putin', 'The Union of Medical Relief Organizations, a US-based charity that works with Syrian hospitals, told the BBC the Damascus Rural Specialty Hospital had confirmed 70 deaths'}



Gambar 5.11 Tampilan Hasil

5.2 Pengujian Perangkat Lunak

Berikut ini akan dijelaskan pengujian perangkat lunak penyelesaian peringkasan multi-dokumen dengan algoritma EM yang dibagi menjadi 2 kasus, yaitu *Summarize with String* dengan data masukan berupa string hasil copy-paste 4 artikel berita tentang *chemical attack* di Douma, dan *Summarize with Links* dengan data masukan berupa *link* sumber dari 4 artikel berita tentang pembunuhan Kim Wall .

5.2.1 Summarize with String

Sebelum masuk ke tahap clustering, dilakukan *pre-processing* terhadap data masukan. Pada kasus ini, 3 tahap *pre-processing* dilakukan yaitu pemisahan kalimat (*split sentence*), tokenisasi, dan penghapusan stopwords (*stopwords removal*). Setelah melakukan *pre-processing*, kata pada tiap dokumen dikelompokkan menjadi 2 cluster. Nilai awal dari titik tengah untuk masing-masing cluster adalah

$$a = (0.000000, 0.000000, 0.200000, 0.000000)$$

$$b = (0.200000, 0.000000, 0.000000, 0.200000)$$

dari nilai awal tersebut, diperoleh bobot keanggotaan yang kemudian digunakan untuk mencari nilai tengah baru. Proses dilakukan berulang sampai konvergen.

5.2.1.1 Summarize with String dengan Dokumen Berbahasa Inggris

Pada setiap proses diperoleh keanggotaan kata pada tiap cluster yang ada pada tabel A dan inisialisasi titik tengah baru pada Tabel B. Hasil ringkasan merupakan kalimat yang memuat 5 kata dengan bobot tertinggi pada tiap cluster.

Dokumen 1:

A chemical attack in a Syrian rebel-held town has sparked widespread international outrage and revulsion. At least 85 people, including many women and children, were killed in Douma on Saturday, according to a statement by rescue workers and medical staff. Condemnation poured in on Sunday as US President

Donald Trump warned there would be a "big price to pay", while Turkey said: "it is not possible to justify nor accept such attacks for any reason and in any way, shape or form." "Many dead, including women and children, in mindless CHEMICAL attack in Syria," Trump wrote on Twitter, lashing out at Syrian President Bashar al-Assad and his ally, Russia's Vladimir Putin. "President Putin, Russia and Iran are responsible for backing Animal Assad. Big price to pay," he said. Trump's threat came exactly a year and a day after the US army fired cruise missiles at a Syrian airbase in retaliation for a deadly sarin gas attack on the rebel-held town of Khan Sheikhoun. Tom Bossert, White House homeland security adviser, told ABC television that he "wouldn't take anything off the table" when asked whether the US could again respond with a missile attack. The Assad government and Russia both denied any use of chemical weapons as "fabrications". The Russian foreign ministry called the latest reports a "provocation", warning against "military intervention under far-fetched and fabricated pretexts". Commenting on the incidents, the European Union called for an international response to the attack. "The evidence points towards yet another chemical attack by the regime," the bloc said in a statement. "It is a matter of grave concern that chemical weapons continue to be used, especially on civilians. The European Union condemns in the strongest terms the use of chemical weapons and calls for an immediate response by the international community." The EU also called on the UN Security Council to re-establish its checks to identify the perpetrators of chemical attacks and on Russia and Iran - the Syrian government's closest allies - to use their influence with al-Assad to prevent further attacks. Meanwhile, French Foreign Minister Jean-Yves Le Drian said France strongly condemned

attacks and bombings by Syrian government forces in the last 24 hours in Douma, adding they were a "gross violation of international humanitarian law. He called for the United Nations Security Council to meet quickly to examine the situation saying that France would work with allies to verify reports that chemical weapons were used. Referring to President Emmanuel Macron's warning that France could strike unilaterally if there was a deadly chemical attack, Le Drian said that Paris would assume all its responsibilities in the fight against the proliferation of chemical weapons. Syria's White Helmets, who are the first responders in rebel-held areas of Syria, said the attack late on Saturday involved "poisonous chlorine gas". In recent years, the Syrian government has been accused of using chemical weapons as a tool against the armed opposition.

Dokumen 2:

US President Donald Trump has lashed out at Syrian President Bashar al-Assad and his allies Russia and Iran over a suspected chemical attack, saying there will be a "big price to pay". Medical sources say dozens of people were killed in an attack on the rebel-held town of Douma on Saturday. The UN Security Council is expected to discuss the crisis on Monday. Both Syria and Russia deny a chemical attack took place and have reached an evacuation deal with the rebels. Outrage at the alleged use of chemical weapons has been widespread with Pope Francis saying, "Nothing, nothing can justify the use of such devices of extermination against defenceless people and populations." The European Union has called for an "immediate response by the international community". One video, recorded by rescue workers known as the White Helmets, shows a number of men, women and children lying lifeless

inside a house, many with foam at their mouths. Other unverified footage shows young children crying as they are treated in a makeshift medical unit. However, it has not been possible to verify independently what actually happened, or the actual number of dead. The Union of Medical Care and Relief Organizations, which runs medical facilities in the Eastern Ghouta, told BBC News that 70 deaths had been confirmed. The reported chemical weapons attack on the rebel-held enclave of Douma poses some hard questions. For the civilians, the families and children there, can they continue to bear the terror, bloodshed and inhumanity being rained upon them? For the militant rebel fighters, do they believe their struggle and continued resistance is still worth the price they and the civilians are being forced to pay by the Syrian government? For Russia, how far is it prepared to defend its allies' apparent use of chemical weapons when its own apparent use of a nerve agent in the UK is subject to so much global condemnation? And for the international community, how will it respond to what is clearly a test of its resolve one year on from the Khan Sheikhoun chemical attack? Will the ritual condemnations be backed by military action in an attempt to deter President Assad's further use of chemical weapons? Or will the West believe that any escalation would simply prolong a conflict that many just want to end?

Dokumen 3:

SHOCKING images of Syrian children left fighting for their lives have beamed around the world following an apparent gas attack on Saturday evening. More than 500 civilians — mostly women and children — are said to be injured following the attack in Douma, with many of those seen at medical centres having difficulty

breathing. Local reports suggest up to 70 people may have been killed, including families found in their homes and shelters. The horrific photos speak for themselves. Be warned — they're graphic. Disturbing footage has also emerged of lifeless men, women and children sprawled on floors and in stairwells, many with white foam coming from their mouths and nostrils. Among the distressing clips was a father begging his dead daughter to wake up and children being hosed down in a hospital. The opposition's Syrian Civil Defence rescue organisation said the victims appeared to have suffocated. They did not identify the substance used, but the civil defence organisation, also known as the White Helmets, and the Syrian American Medical Society, a medical relief organisation, said survivors treated at clinics smelled strongly of chlorine. Those reports could not be independently verified because of a government blockade around the town. Emergency workers said families had been hiding in their basements following earlier air strikes and the gas seeped through the buildings and suffocated them. Australia's Foreign Affairs Minister Julie Bishop issued a statement on Monday condemning the reported attack, and called on parties involved in the continuing conflict in Syria, to "apply maximum pressure on the Syrian regime to fully halt any and all chemical weapons related activity". "The use of chemical weapons at anytime, anywhere, under any circumstances is abhorrent and deserves investigations and accountability," Ms Bishop said. The alleged attack came almost exactly a year after another Syrian chemical attack on civilians, which prompted US President Donald Trump to launch a missile strike on a Syrian air base. Syrian state media today reported a missile strike on a military airport in the country's centre, shortly after Mr Trump issued a new warning to

Damascus and its allies condemning the chemical attack. He warned there will be a "big price to pay" following the "mindless" attack. He took to Twitter to slam the Syrian President Bashar al-Assad as an "animal", while accusing Russia and Iran of backing him.

Dokumen 4:

At least 70 people have died in a suspected chemical attack in Douma, the last rebel-held town in Syria's Eastern Ghouta, rescuers and medics say. Volunteer rescue force the White Helmets tweeted graphic images showing several bodies in basements. It said the deaths were likely to rise. There has been no independent verification of the reports. Syria has called the allegations of a chemical attack a "fabrication" - as has its main ally, Russia. The US state department said Russia - with its "unwavering support" for Syria's government - "ultimately bears responsibility" for the alleged attacks. Several medical, monitoring and activist groups reported details of a chemical attack. "Seventy people suffocated to death and hundreds are still suffocating," said Raed al-Saleh, head of the White Helmets. An earlier, now deleted tweet, put the number dead at more than 150. The pro-opposition Ghouta Media Center tweeted that more than 75 people had "suffocated", while a further 1,000 people had suffered the effects of the alleged attack. It blamed a barrel bomb allegedly dropped by a helicopter which it said contained Sarin, a toxic nerve agent. The Union of Medical Relief Organizations, a US-based charity that works with Syrian hospitals, told the BBC the Damascus Rural Specialty Hospital had confirmed 70 deaths. A spokeswoman said there were reports of people being treated for symptoms including convulsions and

foaming of the mouth, consistent with nerve or mixed nerve and chlorine gas exposure. Continued shelling overnight and on Sunday was making it impossible to reach victims. As the allegations emerged, Syria's state news agency Sana said the reports were invented by the Jaish al-Islam rebels who remain in control in Douma. "Jaish al-Islam terrorists are in a state of collapse and their media outlets are [making] chemical attack fabrications in an exposed and failed attempt to obstruct advances by the Syrian Arab army," Sana said. In August 2013, rockets containing the nerve agent Sarin were fired at rebel-held areas of the Eastern Ghouta, killing hundreds of people. A UN mission confirmed the use of Sarin, but it was not asked to state who was responsible. Western powers said only Syrian government forces could have carried out the attack. In April 2017, more than 80 people died in a Sarin attack on the opposition-held town of Khan Sheikhoun, and a joint inquiry by the UN and the Organization for the Prohibition of Chemical Weapons (OPCW) held the Syrian government responsible. Douma is the last rebel-held town in Syria's Eastern Ghouta region, and is under siege from Russian-backed Syrian government forces. An intense aerial and ground assault launched on Friday, after talks between Moscow and the rebels broke down, is said to have killed dozens of people. Before negotiations failed, Jaish al-Islam had been trying to secure a deal that would let its members stay in Douma as a local security force.

Keterangan Dokumen Sumber:

Dokumen 1: Menjelaskan tentang serangan kimia yang terjadi di Douma dan banyak korban terbunuh dalam serangan tersebut. Presiden Amerika mengecam (di *twitter*) presiden Rusia dan presiden Suriah ada di balik serangan kimia tersebut. Presiden

Amerika mengancam bahwa Amerika bisa saja membalas kasus serangan kimia mengingat sebelumnya pada kasus serangan gas sarin di Khan Sheikhoun, Amerika telah meluncurkan rudal ke *airbase* Suriah sebagai balasan atas kasus serangan gas Sarin tersebut. Di sisi lain, Suriah dan Rusia menolak tuduhan atas serangan kimia yang terjadi. EU mengusulkan agar PBB memperketat keamanan sementara Menteri luar negeri Prancis mengecam keras serangan kimia dan menyatakan ke PBB bahwa Prancis siap untuk ikut memerangi orang-orang dibalik serangan kimia tersebut.

Dokumen 2: Menjelaskan tentang presiden Amerika yang mengecam keras presiden Suriah dan sekutunya, Rusia dan Iran, terhadap dugaan serangan kimia yang terjadi di Douma. Baik Rusia dan Suriah menolak tuduhan serangan kimia tersebut. Rekaman Video yang diambil oleh tim penyelamat *White Helmets* memperlihatkan banyaknya korban yang berjatuh akibat serangan kimia yang terjadi. Artikel ini mempertanyakan adakah keuntungan yang di dapat oleh oknum yang terlibat atas kejadian ini.

Dokumen 3: Menjelaskan tentang banyaknya korban yang berjatuh akibat serangan kimia di Douma dan pernyataan menteri luar negeri Australia yang mengecam adanya serangan kimia yang terjadi.

Dokumen 4: Menjelaskan tentang lebih dari 150 orang tewas dan 1000 orang menderita akibat serangan kimia yang terjadi dan menceritakan tentang serangan gas sarin yang terjadi di Ghouta pada Agustus 2013.

Hasil clustering:

Cluster 1	Cluster 2
['ultimately' 'fired' 'ground' 'convulsions' 'killing' '70' 'al' 'april' 'activist' 'deal' 'deleted' 'helicopter' '2013' 'moscow' 'force' 'organizations' 'exposure']	['ms' '500' 'continuing' 'intervention' 'daughter' 'house' 'suggest' 'discuss' 'smelled' 'may' 'far' 'society' 'condemnation' 'strike']

'confirmed'	'details'	'works'	'following'	'inside'	'first'
'outlets'	'union'	'failed'	'verified'	'television'	'abc'
'unwavering'	'syria'	'helmets'	'adding'	'launch'	'gross'
'media'	'islam'	'chlorine'	'images'	'investigations'	'immediate'
'mixed'	'mission'	'foaming'	'bomb'	'also'	'julie'
'treated'	'rural'	'independent'	'responders'	'activity'	'price'
'russian'	'mouth'	'friday'	'sana'	'anywhere'	'poses'
'hospital'	'town'	'department'	'referring'	'pressure'	'prepared'
'deaths'	'000'	'dropped'	'allegedly'	'nothing'	'use'
'carried'	'say'	'symptoms'	'adviser'	'killed'	'medical'
'responsible'	'trying'	'verification'	'mindless'	'centre'	'world'
'responsibility'	'nerve'	'raed'	'rebel'	'turkey'	'including'
'fabrication'	'opcw'	'held'	'saleh'	'us'	'west'
'white'	'put'	'consistent'	'areas'	'forced'	'pretexts'
'dozens'	'suffered'	'blamed'	'asked'	'widespread'	'substance'
'reach'	'specialty'	'graphic'	'justify'	'response'	'form'
'members'	'number'	'russia'	'allies'	'global'	'children'
'backed'	'said'	'everal'	'barrel'	'families'	'tool'
'agent'	'region'	'talks'	'hospitals'	'1'	'abhorrent'
'least'	'eastern'	'organization'	'bashar'	'wake'	'matter'
'negotiations'	'groups'	'bodies'	'let'	'weapons'	'macron'
'shelling'	'assault'	'reports'	'identify'	'breathing'	'deter'
'obstruct'	'2017'	'emerged'	'bombings'	'ritual'	'military'
'prohibition'	'rescuers'	'jaish'	'situation'	'slam'	'evidence'
'basements'	'several'	'containing'	'men'	'hosed'	'emmanuel'
'ghouta'	'attack'	'opposition'	'siege'	'armed'	'disturbing'
'aerial'	'control'	'effects'	'alleged'	'nations'	'examine'
'exposed'	'seventy'	'contained'	'possible'	'bishop'	'crying'
'remain'	'damascus'	'sunday'	'hiding'	'upon'	'circumstances'
'suffocating'	'rockets'	'tweet'	'un'	'chemical'	'another'
'likely'	'powers'	'joint'	'volunteer'	'issued'	'hard'
'center'	'overnight'	'allegations'	'extermination'	'year'	'twitter'
'toxic'	'secure'	'people'	'launched'	'statement'	'years'
'main'	'charity'	'rebels'	'hundreds'	'according'	'work'
'khan'	'suspected'	'victims'	'makeshift'		'sparked'

'fabrications' 'rise' 'making' 'sarin'	'happened' 'concern'
'broke' 'death' 'showing' 'attempt'	'questions' 'donald' 'day'
'last' 'spokeswoman' 'earlier'	'putin' 'assad' 'using' 'outrage'
'based' 'august' 'bbc' 'ally' 'relief'	'place' 'wrote' '24' 'fight'
'state' 'suffocated' 'terrorists' 'local'	'parties' 'reason' 'retaliation'
'still' 'pro' 'impossible' 'head' '75'	'reported' 'violation' 'accept'
'forces' 'bears' 'agency' 'continued'	'table' 'saturday' 'around'
'news' 'western' 'intense' 'arab'	'continue' 'footage' 'denied'
'inquiry' 'tweeted' 'would'	'shocking' 'airport' 'rescue'
'government' 'sheikhoun' 'security'	'strikes' 'civilians' 'appeared'
'support' 'collapse' '150' 'told'	'defend' 'new' 'sources'
'advances' 'army' '80' 'monitoring'	'centres' 'halt' 'accountability'
'medics' 'died' 'invented' 'stay'	'american' 'french' 'country'
'douma']	'apply' 'vladimir' 'francis'
	'workers' 'deny' 'fighters'
	'cruise' 'women' 'poured'
	'towards' 'father' 'bossert'
	'united' 'incidents'
	'organisation' 'stairwells'
	'fabricated' 'homeland'
	'whether' 'worth' 'council'
	'independently' 'strongest'
	'staff' 'among' 'took' 'horrific'
	'bloodshed' 'mouths' 'clinics'
	'way' 'establish'
	'condemnations' 'revulsion'
	'minister' 'law' 'airbase' 'air'
	'beamed' 'shelters' 'clearly'
	'missile' 'reached' 'tom' 'yet'
	'injured' 'international'
	'known' 'attacks' 'warned'
	'influence' 'inhumanity'
	'prompted' 'conflict' 'eu'
	'prevent' 'proliferation' 'clips'
	'seeped' 'points' 'saying'

	'trump' 'today' 'found' 'young' 'unverified' 'emergency' 'however' 'lifeless' 'fetched' 'difficulty' 'checks' 'actually' 'australia' 'recent' 'called' 'defence' 'struggle' 'accused' 'latest' 'foreign' 'assume' 'actual' 'backing' 'one' 'jean' 'pope' 'buildings' 'coming' 'much' 'terror' 'quickly' 'populations' 'shows' 'believe' 'fighting' 'enclave' 'involved' 'pay' 'video' 'gas' '85' 'subject' 'yves' 'speak' 'uk' 'evacuation' 'distressing' 'regime' 'bloc' 'simply' 'foam' 'lives' 'meanwhile' 'dead' 'bear' 'exactly' 'missiles' 'strongly' 'anything' 'escalation' 'grave' 'poisonous' 'evening' 'could' 'begging' 'used' 'anytime' 'responsibilities' 'related' 'lying' 'shortly' 'drian' 'verify' 'ministry' 'blockade' 'meet' 'big' 'maximum' 'respond' 'test' 'many' 'civil' 'provocation' 'end' 'shape' 'apparent' 'commenting' 'condemned' 'terms' 'european' 'take' 'nostrils' 'almost' 'paris' 'le' 'devices' 'resistance' 'animal' 'deadly' 'especially' 'survivors' 'mr' 'calls' 'resolve' 'facilities' 'unit' 'unilaterally' 'mostly' 'deserves' 'humanitarian' 'late'
--	--

	'runs' 'perpetrators' 'community' 'syrian' 'condemning' 'president' 'threat' 'monday' 'lashing' 'prolong' 'lashed' 'recorded' 'militant' 'homes' 'floors' 'condemns' 'defenceless']
--	--

Berikut adalah 5 kata dengan bobot tertinggi pada masing-masing cluster

Cluster 1: ['would' 'works' 'white' 'western' 'volunteer']

Cluster 2: ['yves' 'young' 'yet' 'years' 'year']

Hasil ringkasan:

{ " Referring to President Emmanuel Macron's warning that France could strike unilaterally if there was a deadly chemical attack, Le Drian said that Paris would assume all its responsibilities in the fight against the proliferation of chemical weapons", " For the civilians, the families and children there, can they continue to bear the terror, bloodshed and inhumanity being rained upon them? For the militant rebel fighters, do they believe their struggle and continued resistance is still worth the price they and the civilians are being forced to pay by the Syrian government? For Russia, how far is it prepared to defend its allies' apparent use of chemical weapons when its own apparent use of a nerve agent in the UK is subject to so much global condemnation? And for the international community, how will it respond to what is clearly a test of its resolve one year on from the Khan Sheikhou chemical attack? Will the ritual condemnations be backed by military action in an attempt to deter President Assad's further use of chemical weapons? Or will the West believe that any escalation would simply prolong a conflict that many just want to end?", ' In recent years, the Syrian government has been accused of using chemical weapons as a

tool against the armed opposition', ' Before negotiations failed, Jaish al-Islam had been trying to secure a deal that would let its members stay in Douma as a local security force', ' Other unverified footage shows young children crying as they are treated in a makeshift medical unit', ' Condemnation poured in on Sunday as US President Donald Trump warned there would be a "big price to pay", while Turkey said: "it is not possible to justify nor accept such attacks for any reason and in any way, shape or form", ' The alleged attack came almost exactly a year after another Syrian chemical attack on civilians, which prompted US President Donald Trump to launch a missile strike on a Syrian air base', ' The Union of Medical Relief Organizations, a US-based charity that works with Syrian hospitals, told the BBC the Damascus Rural Specialty Hospital had confirmed 70 deaths', ' Disturbing footage has also emerged of lifeless men, women and children sprawled on floors and in stairwells, many with white foam coming from their mouths and nostrils', " Trump's threat came exactly a year and a day after the US army fired cruise missiles at a Syrian airbase in retaliation for a deadly sarin gas attack on the rebel-held town of Khan Sheikhoun", ' "The evidence points towards yet another chemical attack by the regime," the bloc said in a statement', 'Tom Bossert, White House homeland security adviser, told ABC television that he "wouldn't take anything off the table" when asked whether the US could again respond with a missile attack', ' He called for the United Nations Security Council to meet quickly to examine the situation saying that France would work with allies to verify reports that chemical weapons were used'}

Keterangan Hasil Ringkasan:

Hasil ringkasan menjelaskan tentang pernyataan Prancis yang siap ikut serta memerangi pelaku serangan kimia (Dokumen 1).

Kemudian mempertanyakan keuntungan apa yang di dapat oleh pihak yang terkait dengan peristiwa ini (Dokumen 2). Hasil ringkasan juga mendeskripsikan tentang korban serangan kimia yang terjadi di Douma (Semua dokumen membahas korban pada tragedi ini) dan pernyataan presiden Amerika atas tuduhnya terhadap presiden Suriah dan presiden Rusia atas terjadinya serangan kimia ini (Dokumen 1 dan dokumen 2).

5.2.1.2 Summarize with String dengan Dokumen Masukan Berbahasa Indonesia

Dokumen 1:

Khalayak luas telah merayakan dengan sukacita kelahiran anak ketiga Pangeran William dan Kate Middleton, yang diberi nama Pangeran Louis Arthur Charles. Pada hari Jumat, 27 April 2018, Kensington Palace mengumumkan bahwa anggota baru keluarga Kerajaan Inggris itu kelak diberi gelar His Royal Highness Prince Louis of Cambridge. Dikutip dari Time.com pada Minggu (29/3/2018), muncul perbincangan tentang latar belakang penamaan Louis, yang disebut memiliki latar belakang kisah yang menyedihkan. Kemungkinan besar pemilihan nama bayi tersebut terinspirasi oleh Lord Louis Mountbatten, nama paman dari Pangeran Charles yang dibunuh oleh Tentara Republik Irlandia (IRA) pada 1979. Louis Mountbatten--cicit dari Ratu Victoria--adalah seorang perwira Angkatan Laut Kerajaan Inggris yang pernah bertempur di Perang Dunia I dan Perang Dunia II. Ia diketahui menjadi kapten kapal perusak H.M.S. Kelly yang tenggelam di lepas pantai Jerman, akibat diserang torpedo Nazi pada 1941. Pascaserangan tersebut, ia dan sebagian awaknya yang selamat, berikrar akan tetap berada di garis terdepan untuk melindungi Inggris. Atas kegigihannya, majalah TIME sempat memuat foto

wajahnya di halaman sampul salah satu edisi yang terbit pada 1942 silam. Setahun kemudian, ia justru dikecam karena memimpin serangan brutal ke Kota Dieppe di Prancis utara, yang mencoreng wibawanya sebagai pemimpin. Meski begitu, Perdana Menteri Winston Churchill tetap menunjuknya untuk memimpin invasi D-Day yang bersejarah pada Juni 1944. Louis Mountbatten meninggal pada usia 79, ketika IRA meledakkan bom yang terpasang di kapal layarnya pada Agustus 1979. Tiga orang lainnya turut tewas dalam serangan itu, termasuk cucunya yang berusia 14 tahun, Nicholas Knatchbull. "Baik ketika memimpin operasi perusak di tengah pertempuran Perang Dunia II, atau kemudian mewakili Inggris dalam gelombang kemerdekaan India, Mountbatten telah mengakumulasi kemenangan publik yang tampaknya ajaib," tulis TIME dalam obituari sang perwira angkatan laut. Sosok Lord Mountbatten baru-baru ini kembali dihidupkan dalam serial televisi Netflix, *The Crown*, yang diperankan oleh aktor Inggris, Greg Wise. Menariknya, nama Louis juga memiliki keterkaitan lain dalam sejarah Kerajaan Inggris. Kakek sang bayi, Pangeran Phillip, merupakan anak dari bangsawan Denmark bergelar Prince Louis from Battenberg.

Dokumen 2:

Memberi nama seorang anak bukan hal yang selalu mudah bagi sebagian besar orangtua, termasuk bagi Kate Middleton dan Pangeran William. Apalagi, semua nama bayi yang lahir dari keluarga Kerajaan Inggris memiliki makna sejarah yang mendalam. Hal ini pun berlaku bagi anak ketiga Kate dan William yang bernama Louis Arthur Charles. Hari jumat (27/4/2018) lalu, Istana Kensington telah mengumumkan nama anak

ketiga dari pasangan tersebut. Anak tertua Pangeran William dan Kate, Pangeran George, kemungkinan besar akan menjadi raja ketujuh dalam dinasti Kerajaan Inggris. Nama George mengikuti ayah Ratu Elizabeth, Raja George VI, dan kakeknya, Raja George V. Sementara nama Putri Charlotte, inspirasi namanya kemungkinan besar diambil dari kakeknya, Pangeran Charles, dan nama tengahnya diambil dari saudara perempuan Kate, Pippa Matthews. Lantas, bagaimana dengan anak ketiga Pangeran William dan Kate Middleton? Berdasarkan laporan laman Independent, nama Louis sering dikaitkan dengan monarki Perancis, dengan Raja Perancis terakhir bernama Louis XVI yang memerintah dari tahun 1774-1791. Namun, nama itu juga memiliki banyak koneksi dengan keluarga Pangeran Philip atau the Duke of Edinburgh. Pangeran Philip memiliki seorang paman dan guru pribadi bernama Lord Louis Mountbatten. Pangeran William dan anak pertamanya, Pangeran George, juga memiliki nama yang sama. Louis terpilih sebagai nama tengah ketiga Pangeran William ketika ia lahir, sedangkan nama lengkap Pangeran George adalah George Alexander Louis. Bukan hanya nama depan bayi kerajaan yang telah menarik perhatian. Pada satu titik, Arthur dianggap sebagai nama paling mungkin untuk anggota baru keluarga Kerajaan Inggris ini. Namun, nama tersebut ternyata dipilih sebagai nama tengah bayi tersebut. Sementara, Charles yang menjadi nama belakang, jelas merujuk pada kakeknya yang bernama Pangeran Charles. Raja Arthur yang legendaris jelas muncul dalam pikiran kita ketika mendengar nama Arthur.

Berikut adalah 5 kata dengan bobot tertinggi pada masing-masing cluster

Cluster 1: ['xvi' 'william' 'vi' 'v' 'titik']

Cluster 2: ['yang' 'wise' 'winston' 'wibawanya' 'wajahnya']

Hasil Ringkasan:

{ ' Atas kegigihannya, majalah TIME sempat memuat foto wajahnya di halaman sampul salah satu edisi yang terbit pada 1942 silam', ' Lantas, bagaimana dengan anak ketiga Pangeran William dan Kate Middleton? Berdasarkan laporan laman Independent, nama Louis sering dikaitkan dengan monarki Perancis, dengan Raja Perancis terakhir bernama Louis XVI yang memerintah dari tahun 1774-1791', 'Khalayak luas telah merayakan dengan sukacita kelahiran anak ketiga Pangeran William dan Kate Middleton, yang diberi nama Pangeran Louis Arthur Charles', ' "Baik ketika memimpin operasi perusak di tengah pertempuran Perang Dunia II, atau kemudian mewakili Inggris dalam gelombang kemerdekaan India, Mountbatten telah mengakumulasi kemenangan publik yang tampaknya ajaib," tulis TIME dalam obituari sang perwira angkatan laut', ' Raja Arthur yang legendaris jelas muncul dalam pikiran kita ketika mendengar nama Arthur', ' Pangeran William dan anak pertamanya, Pangeran George, juga memiliki nama yang sama', ' Sementara, Charles yang menjadi nama belakang, jelas merujuk pada kakeknya yang bernama Pangeran Charles', 'com pada Minggu (29/3/2018), muncul perbincangan tentang latar belakang penamaan Louis, yang disebut memiliki latar belakang kisah yang menyedihkan', ' Meski begitu, Perdana Menteri Winston Churchill tetap menunjuknya untuk memimpin invasi D-Day yang bersejarah pada Juni 1944', ' Setahun kemudian, ia justru dikecam karena memimpin serangan brutal ke Kota Dieppe di Prancis

utara, yang mencoreng wibawanya sebagai pemimpin', 'Kemungkinan besar pemilihan nama bayi tersebut terinspirasi oleh Lord Louis Mountbatten, nama paman dari Pangeran Charles yang dibunuh oleh Tentara Republik Irlandia (IRA) pada 1979', ' Pada satu titik, Arthur dianggap sebagai nama paling mungkin untuk anggota baru keluarga Kerajaan Inggris ini', ' Bukan hanya nama depan bayi kerajaan yang telah menarik perhatian', ' Sosok Lord Mountbatten baru-baru ini kembali dihidupkan dalam serial televisi Netflix, *The Crown*, yang diperankan oleh aktor Inggris, Greg Wise', ' Tiga orang lainnya turut tewas dalam serangan itu, termasuk cucunya yang berusia 14 tahun, Nicholas Knatchbull', 'Memberi nama seorang anak bukan hal yang selalu mudah bagi sebagian besar orangtua, termasuk bagi Kate Middleton dan Pangeran William', ' Kelly yang tenggelam di lepas pantai Jerman, akibat diserang torpedo Nazi pada 1941', ' Pascaserangan tersebut, ia dan sebagian awaknya yang selamat, berikrar akan tetap berada di garis terdepan untuk melindungi Inggris', ' Louis Mountbatten meninggal pada usia 79, ketika IRA meledakkan bom yang terpasang di kapal layarnya pada Agustus 1979', ' Hal ini pun berlaku bagi anak ketiga Kate dan William yang bernama Louis Arthur Charles', ' Apalagi, semua nama bayi yang lahir dari keluarga Kerajaan Inggris memiliki makna sejarah yang mendalam', ' Louis Mountbatten--cicit dari Ratu Victoria--adalah seorang perwira Angkatan Laut Kerajaan Inggris yang pernah bertempur di Perang Dunia I dan Perang Dunia II'}

5.2.2 Summarize with Links

Sebelum masuk ke tahap clustering, dilakukan *pre-processing* terhadap data masukan. Pada kasus ini, 6 tahap *pre-processing* dilakukan yaitu *download* artikel dari link masukan,

parsing artikel, pengubahan format artikel dari web ke bentuk text, pemisahan kalimat (*split sentence*), tokenisasi, dan penghapusan stopwords (*stopwords removal*). Setelah melakukan *pre-processing*, kata pada tiap dokumen dikelompokkan menjadi 2 cluster. Nilai awal dari titik tengah untuk masing-masing cluster adalah

$a = (0.000000, 0.000000, 0.200000, 0.000000)$

$b = (0.200000, 0.000000, 0.000000, 0.200000)$

dari nilai awal tersebut, diperoleh bobot keanggotaan yang kemudian digunakan untuk mencari nilai tengah baru. Proses dilakukan berulang sampai konvergen. Pada setiap proses diperoleh keanggotaan kata pada tiap cluster yang ada pada Tabel C dan inisialisasi titik tengah baru pada Tabel D. Hasil ringkasan merupakan kalimat yang memuat 5 kata dengan bobot tertinggi pada tiap cluster.

Link 1 = "<http://www.foxnews.com/world/2018/04/25/kim-wall-killer-peter-madsen-sentenced-to-life-in-prison.html>"

Link 2 = "<http://www.oxygen.com/crime-time/kim-wall-accused-killer-peter-madsen-psychopath-exam>"

Link 3 = "<https://www.cbsnews.com/news/kim-wall-peter-madsen-murder-trial-verdict-denmark-live-updates-2018-4-25/>"

Link 4 = "<http://www.bbc.com/news/world-europe-43894495>"

Tabel 5.1 adalah hasil clustering pada proses peringkasan 4 dokumen masukan menggunakan *link*.

Tabel 5.1 Hasil Clustering

Cluster 1	Cluster 2
['ms' '40' '6' 'tortured'	['coast' 'motivated' 'plastic'
'formidable' 'driven' 'far'	'may' 'deserve' 'burkoe'
'london' 'unsupported'	'offer' 'argued' 'following'
'next' 'story' 'return' '1987'	'inside' 'aboard' 'first'
'journalist' 'also' 'anna'	'court' 'equates' 'empathy'
'sentenced' 'studied'	'freelance' 'city'
'oxbridge' 'danish' 'south'	'emotionally' 'hald'

'legs'	'raised'	'overboard'	'outside'	'jepsen'	'taken'
'films'	'born'	'koge'	'ended'	'crimes'	'given'
'missing'	'prosecutors'		'finally'	'activity'	
'50km'	'murder'	'kim'	'necessary'	'divide'	
'media'	'carbon'	'images'	'removed'	'details'	
'30'	'hit'	'47'	'pressure'	'indecent'	
'seen'	'friends'	'weighted'	'ordered'	'help'	
'town'	'homemade'	'rado'	'undocumented'	'crime'	
'bags'	'honours'	'trelleborg'	'appear'	'personal'	'killed'
'dropped'	'throat'	'planned'	'aid'	'defense'	'including'
'character'	'tendencies'		'danger'	'clear'	'exhibited'
'say'	'10'	'sea'	'semen'	'ap'	'accident'
'discovered'	'claim'		'corpse'	'case'	'madsen'
'previously'	'school'		'jurors'	'sought'	'woman'
'venture'	'within'	'year'	'horrible'	'various'	'manner'
'boyfriend'	'dividing'		'guilt'	'written'	'metal'
'journalism'	'terrible'		'team'	'offering'	
'october'	'happened'		'committed'	'read'	
'masters'	'arrest'	'bay'	'extended'	'press'	'never'
'premeditated'	'place'		'private'	'deaths'	'pleaded'
'telling'	'classmate'	'cyclist'	'peter'	'motive'	'quoted'
'top'	'close'	'economics'	'severe'	'evidence'	'uc3'
'account'	'events'	'opening'	'via'	'baltic'	'hours'
'denied'	'university'	'board'	'make'	'early'	'conviction'
'interviewed'			'12'	'piece'	'accepted'
'unconvincing'			'8'	'respect'	'wanted'
'researching'	'new'	'police'	'beijing'	'bury'	'interview'
'departed'	'suspicion'		'torture'	'added'	'established'
'women'	'codrea'		'explanations'	'multiple'	
'admitted'	'accidental'		'sentence'	'explanation'	
'either'	'family'	'getty'	'fell'	'credible'	'years'
'playback'	'vessel'		'parts'	'wonderful'	'according'
'throwing'	'somewhere'		'photos'	'detailed'	'genitals'
'monoxide'	'clothing'	'11'	'magazine'	'prior'	'betina'
'winning'	'sweden'		'problem'	'set'	'lawyer'

'swedish'	'assault'	'excerpts'	'consistent'
'psychopathic'	'device'	'something'	'moved'
'computer'	'parole'	'quietly'	'verdict'
'remains'	'differing'	'suffocate'	'burying'
'changed'	'even'	'afp'	'publications'
'month'	'gave'	'mutilated'	'bax'
'sank'	'international'	'16'	'sexual'
'session'	'programme'	'arguing'	'report'
'waters'	'copyright'	'randomly'	'develops'
'deliberately'	'spotted'	'remorse'	'said'
'convict'	'suffocating'	'showed'	'spacecraft'
'dismembering'	'helped'	'traits'	'region'
'accounts'	'body'	'brutal'	'bed'
'investigators'	'columbia'	'described'	'decision'
'boarding'	'since'	'featured'	'lindhardt'
'southern'	'believe'	'sexually'	'decided'
'disappeared'	'class'	'go'	'whether'
'across'	'know'	'trial'	'explaining'
'22'	'150lb'	'island'	'along'
'cohort'	'strait'	'took'	'spare'
'submarine'	'evening'	'timeline'	'judge'
'passing'	'august'	'stabbed'	'determined'
'friend'	'tonne'	'2017'	'acted'
'placed'	'poisoning'	'emotion'	'conceal'
'went'	'days'	'agreed'	'confessed'
'miles'	'admitting'	'nordic'	'prosecutor'
'dismissed'	'rescued'	'dismember'	'several'
'without'	'small'	'caused'	'reached'
'self'	'alarm'	'stopped'	'psychiatric'
'initially'	'caption'	'aug'	'underwear'
'gain'	'community'	'testifying'	'authorities'
'copenhagen'	'wall'	'ride'	'tried'
'image'	'divers']	'saying'	'found'
		'horror'	'restrictions'
		'offered'	'charged'
		'lifeless'	'buch'
		'march'	'impaired'
		'anger'	'conducted'
		'good'	'made'
		'impaled'	'denmark'
		'arms'	'upwards'
		'sailing'	'trustworthy'
		'sky'	'lack'
		'cause'	'pitch'
		'allegations'	'toxic'
		'sudden'	'manned'
		'pieces'	'accused'
		'old'	

	'videos' 'open' 'associated' 'nautilus' 'one' 'abroad' 'co' 'immediately' 'walls' 'shows' 'heard' 'accidentally' 'anette' 'prison' 'york' 'gruesome' 'death' 'special' 'life' 'smaller' 'hearing' 'wounds' 'guardian' 'departing' 'chosen' 'large' 'lift' 'tools' 'give' 'reuters' 'dead' 'risk' 'tied' 'stockings' 'guilty' 'could' 'sink' 'jakob' 'times' 'dismembered' 'based' 'opinion' 'founder' 'weeks' 'brought' 'big' 'gory' 'claims' 'appeal' 'prosecution' 'claimed' 'stab' 'builds' 'narcissistic' 'impossible' 'broadcasting' 'buildup' 'began' 'aloud' 'take' 'torturing' 'torso' 'hatches' 'fumes' 'engmark' 'news' 'psychiatrists' 'especially' 'cited' 'would' 'inventor' 'contributed' 'dismemberment' 'assessment' 'repeatedly' 'normally' 'stood' 'wednesday' 'photograph' 'trip' 'company' 'two' 'result' 'scanpix' 'died' 'claiming' 'serious' 'ship' 'insane' 'unable' 'others' 'incident' 'thursday' 'stay'
--	--

	'listening' 'murdering' 'cutting']	'reportedly' 'wired'
--	--	-------------------------

Berikut adalah 5 kata dengan bobot tertinggi pada masing-masing cluster

Cluster 1: ['year' 'women' 'without' 'within' 'winning']

Cluster 2: ['york' 'years' 'written' 'wounds' 'would']

Hasil Ringkasan:

{ " KIM WALL'S GRUESOME DEATH DETAILED IN PETER MADSEN MURDER TRIAL: A TIMELINE Prosecutor Jakob Buch-Jepsen said during the trial that Wall's dismembered torso had multiple stab wounds, including in the genitals, and that semen had been found in Madsen's underwear", ' Wall had written for various publications including The New York Times and The Guardian, and she planned to pitch her Madsen story to Wired magazine', ' Wall, a 30-year-old freelance journalist, set out on the submarine on Aug', 'Image copyright Getty Images Image caption Madsen was sentenced to life for the murder of Swedish journalist Kim Wall Danish inventor Peter Madsen has been sentenced to life in prison without parole for the murder of the Swedish journalist Kim Wall on his submarine', ' During the opening session of his trial last month, prosecutors said there was a suspicion that he had "psychopathic tendencies" after investigators discovered films on his computer showing women being tortured and mutilated', ' Even within her cohort she was top of the class, winning honours in her year, her classmate and friend Anna Codrea-Rado told the BBC', ' In Denmark, a life sentence equates to 16 years, which can be extended if necessary', " It was not immediately clear whether Madsen's legal team would appeal the verdict", ' She disappeared days before she and her boyfriend would have moved to Beijing', ' Her dismembered

remains were found by Danish police at sea on 21 August last year, 11 days after she interviewed him on board his homemade vessel', ' In Denmark, life equates to 16 years, which can be extended if necessary', ' Buch-Jepsen also said videos of women being tortured and killed were found on his personal computer — videos that he did not make himself}

5.2.3 Validasi Hasil Ringkasan

Pada kasus 5.2.1, perhitungan dan perbandingan jumlah kata pada hasil ringkasan dengan jumlah kata pada dokumen sumber dilakukan untuk membuktikan bahwa sistem mampu menghasilkan ringkasan dengan bentuk yang lebih sederhana dibandingkan dengan dokumen sumber. Perhitungan jumlah kata dapat dilihat pada Tabel 5.2 sebagai berikut

Tabel 5.2 Hasil Perbandingan Jumlah Kata

Topik	Sumber 1	Sumber 2	Sumber 3	Sumber 4	Hasil Ringkasan
Chemical Attack in Douma (English)	484	X	X	X	105
Kelahiran Anak Ketiga Kate & William (Indonesia)	324	X	X	X	274
Royal Baby 3 (English)	288	X	X	X	61
Arsene Wenger (Indonesia)	212	X	X	X	133
Chemical Attack in Douma (English)	484	384	X	X	309

Kelahiran Anak Ketiga Kate & William (Indonesia)	324	276	X	X	430
Royal Baby 3 (English)	491	816	X	X	415
Royal Baby 3 (English)	816	291	491	X	429
Chemical Attack in Douma (English)	484	384	387	475	503

Berdasarkan percobaan terhadap 4 topik berbeda dengan satu dokumen masukan, sistem mampu menurunkan jumlah kata pada hasil ringkasan sehingga hasil ringkasan dapat dinyatakan lebih sederhana. Namun pengujian terhadap hasil ringkasan dengan lebih dari satu dokumen masukan tidak selalu menghasilkan jumlah kata yang lebih sedikit. Hal ini dikarenakan perbedaan informasi penting pada masing-masing sumber. Hasil ringkasan memuat informasi-informasi penting tersebut sehingga mempengaruhi peningkatan jumlah kata apabila dibandingkan dengan tiap-tiap sumber yang ada. Akan tetapi, apabila dibandingkan dengan rata-rata kata yang ada pada dokumen sumber, jumlah kata pada hasil ringkasan lebih sedikit sehingga hasil ringkasan dapat dikatakan lebih sederhana dibandingkan dengan dokumen sumber.

Kemudian akan dihitung persentase keakuratan hasil ringkasan menggunakan similaritas *cosine* pada Persamaan (3.1) dengan nilai similaritas inter dan similaritas intra pada persamaan 3.6 dan 3.13, diperoleh nilai inter dan intra sebagai berikut,

$$\text{inter_sim} = 0.478243452595$$

$$\text{intra_sim} = 0.372755884523$$

Maka hasil untuk pengujian akurasi proses peringkasan 4 dokumen berita dengan *string* sebagai dokumen masukan pada contoh 5.2.1 adalah bahwa hasil ringkasan mampu menaikkan similaritas terhadap dokumen masukan sebanyak 20%.

Pada kasus 5.2.2, akan dihitung persentase keakuratan hasil ringkasan menggunakan rumus di atas, diperoleh nilai inter dan intra sebagai berikut

$$\text{inter_sim} = 0.4200114883$$

$$\text{intra_sim} = 0.511322065922$$

Maka hasil untuk pengujian akurasi proses peringkasan 4 dokumen berita dengan *link* sebagai dokumen masukan pada contoh 5.2.2 adalah bahwa hasil ringkasan menurunkan similaritas terhadap dokumen masukan sebanyak 17%.

Dilakukan pengujian kembali terhadap 4 dokumen berita dengan *link* sebagai dokumen masukan menggunakan topik lain dan didapatkan nilai similaritas inter dan intra pada Tabel 5.1 sebagai berikut.

Tabel 5.3 Hasil Pengujian Nilai Similaritas Intra dan Inter

Topik	inter_sim	intra_sim
<i>Kim Wall's Murder</i>	0.42001148	0.51132206
Kelahiran Anak Ketiga <i>Prince William</i> dan <i>Princess Kate</i>	0.59388639	0.47380150
<i>The Transfer of US Embassy in Israel</i>	0.53116998	0.37509215

<i>The Dead of Palestinean Lecturer in Malaysia</i>	0.38722894	0.25667354
<i>Meghan Markle's Profile</i>	0.50052617	0.26838805
<i>Chemical Attack in Douma</i>	0.52518815	0.51792048
<i>Israeli Soldier was Set Free after Killing Palestinian</i>	0.33467044	0.32405098
<i>Cole Sprouse and Lili Reinhart at the Met Gala</i>	0.41235635	0.32110488
<i>Arsene Wenger Berhenti dari Arsenal</i>	0.47560149	0.33924596
<i>Kim Jong-un to meet Moon Jae- in</i>	0.54003116	0.31968956

Berdasarkan percobaan yang dilakukan terhadap 10 topik berbeda, hasil ringkasan atas 9 dari 10 topik artikel berita menghasilkan nilai similaritas inter yang lebih besar dibandingkan dengan nilai similaritas intra. Hal tersebut menunjukkan bahwa hasil ringkasan yang diperoleh sistem mampu untuk menaikkan nilai similaritas terhadap dokumen masukkan. Oleh karena itu, ketepatan ringkasan dalam memuat informasi yang ada pada dokumen sumber adalah sebesar 90%. Setiap kalimat yang diambil dari artikel-artikel sumber sebagai hasil ringkasan

merepresentasikan informasi independen atas topik yang ada, sehingga kalimat satu dan lainnya pada hasil ringkasan dapat dijamin keterhubungannya selama artikel berita memiliki kesamaan topik.

BAB VI PENUTUP

Pada bab ini berisi tentang beberapa kesimpulan yang dihasilkan berdasarkan penelitian yang telah dilaksanakan dan saran yang dapat digunakan jika penelitian ini dikembangkan.

6.1 Kesimpulan

Berdasarkan percobaan terhadap 4 topik berbeda dengan satu dokumen masukan, sistem mampu menurunkan jumlah kata pada hasil ringkasan sehingga hasil ringkasan dapat dinyatakan lebih sederhana. Kemudian berdasarkan percobaan yang dilakukan terhadap 10 topik berbeda, hasil ringkasan atas 9 dari 10 topik artikel berita menghasilkan nilai similaritas inter yang lebih besar dibandingkan dengan nilai similaritas intra. Hal tersebut menunjukkan bahwa hasil ringkasan yang diperoleh sistem mampu untuk menaikkan nilai similaritas terhadap dokumen masukan. Oleh karena itu, ketepatan ringkasan dalam memuat informasi yang ada pada dokumen sumber adalah sebesar 90%. Hal ini menunjukkan bahwa hasil ringkasan mampu merepresentasikan keseluruhan dokumen dengan bentuk yang lebih sederhana sehingga pengguna tetap akan mendapatkan informasi dari tiap dokumen dengan hanya membaca hasil ringkasan.

6.2 Saran

Hal yang penulis sarankan untuk pengembangan penelitian selanjutnya adalah membuat variasi jumlah cluster pada proses clustering menggunakan algoritma EM, memperbanyak jumlah dokumen yang dapat diringkas, dan pemanggilan hasil ringkasan menggunakan *query*.

DAFTAR PUSTAKA

- [1] NetCraft, September 2017 Web Server Survey, England, <https://news.netcraft.com/archives/2017/09/11/september-2017-web-server-survey.html>
- [2] Lee, J.-H., Park, S., Ahn, C.-M., Kim, D. (2009). "Automatic Generic Document Summarization Based on Non-negative Matrix Factorization". Information Processing and Management, vol. 45, pp. 20–34. Elsevier ISSN 0306-4573.
- [3] García-Hernández, R.A., Montiel, R., Ledeneva, Y., Rendón, E., Gelbukh, A., Cruz, R. (2008). "Text Summarization by Sentence Extraction Using Unsupervised Learning". Gelbukh, A., Morales, E.F. (eds.) MICAI. LNCS (LNAI), vol. 5317, pp. 133–143. Springer, Heidelberg.
- [4] Luhn, H.P. (1958). "The automatic creation of Literature abstracts". IBM Journal of Research and Development.
- [5] Edmondson, H.P. (1969). "New Methods in Automatic Extraction". Journal of the Association for Computing Machinery
- [6] Goldstein, J., Carboell, J., Kantrowitz, M., Mittal, V. (1999). "Summarizing text documents: sentence and evaluation metrics". 22nd Int. ACM SIGIR Research and Development in Information Retrieval, Berkley.
- [7] Yulia Ledeneva, René García Hernández, Romyna Montiel Soto, Rafael Cruz Reyes, and Alexander Gelbukh. (2011). "EM Clustering Algorithm for Automatic Text Summarization". Batyrshin and G. Sidorov (Eds.): MICAI 2011, Part I, LNAI 7094, pp. 305–315.
- [8] Yllias Chali, Shafiq R. Jothy, dan Sadid A. Hasan (2009). "Complex Question Answering: Unsupervised Learning Approaches and Experiments". Journal of Artificial Intelligence Research 35 1-47.

- [9] Ms. Meghana. N.Ingole, Mrs.M.S.Bewoor, Mr.S.H.Patil (2012). "Text Summarization using Expectation Maximization Clustering Algorithm". International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com Vol. 2, Issue 4, July-August, pp.168-171.
- [10] Ziqiang Cao, Wenjie Li, Sujian Li, Furu Wei (2017). "Improving Multi Document Summarization via Text Classification". Association for the Advancement of Artificial Intelligence (www.aaai.org).
- [11] Sparck-Jones, K. (1999). "Automatic summarizing: factors and directions:. Mani, I.; Maybury, M. Advances in Automatic Text Summarization. The MIT Press 1-12.
- [12] Mani, I. (2001). "Automatic Summarization". J.Benjamins Publ. Co. Amsterdam Philadelphia.
- [13] Salton, G. Buckley, C. (1988). "Term-weighting approaches in automatic text retrieval".Information Processing and Management 24, 513-523. Reprinted in: Sparck-Jones, K.; Willet, P. (eds.) Readings in I.Retrieval. Morgan Kaufmann (1997) 323-328.
- [14] Mani, I., House, D., Klein, G., Hirschman, L., Obrsl, L., Firmin, T., Chrzanowski, M., Sundheim, B. (1998). "The TIPSTER SUMMAC Text Summarization Evaluation". MITRE Technical Report MTR 98W0000138. The MITRE Corporation.
- [15] Mitra, M.; Singhal, A.; Buckley, C. (1997) "Automatic text summarization by paragraph extraction". In Proceedings of the ACL'97/EACL'97 Workshop on Intelligent Scalable Text Summarization. Madrid.
- [16] Rath, G. J., Resnick A., Savvage R. (1961) "The formation of abstracts by the selection of sentences". American Documentation 12 (2) 139-141.

- [17] Joel Larocca Neto, Alex A. Freitas, Celso A. A. Kaestner. (2015). "Automatic Text Summarization using a Machine Learning Approach". Curitiba – PR. 80.215-901. BRAZIL.
- [18] Mani, I., Bloedorn. (1998). "E. Machine Learning of Generic and User-Focused Summarization". In Proceedings of the Fifteenth National Conference on AI (AAAI-98) 821-826.
- [19] Ronan Collobert, Jason Weston, Leon Bottou, Michael Karlen, Koray Kavukcuoglu, Pavel Kuksa. (2011). "Natural Language Processing (Almost) from Scratch". Journal of Machine Learning Research 12 2493-2537.
- [20] SAS Institut Inc. "Natural Language Processing. What It Is and Why It Matters". SAS Insight Analytics.
- [21] C. D. Manning, P. Raghavan dan H. Schütze, An Introduction to Information Retrieval, Cambridge, England: Cambridge University Press, 2009.
- [22] Rao AIMSCS, C. R. (2015). "Big Data Analytics". University of Hyderabad Campus, Hyderabad, India.
- [23] Wikipedia. *Euclidean Distance*, https://en.wikipedia.org/wiki/Euclidean_distance, (diakses pada 15 April 2018).
- [24] Aprianto, Kasiful. "Optimasi Kernel K-Means dalam Pengelompokan Kabupaten/Kota Berdasarkan Indeks Pembangunan Manusia di Indonesia". Limits J. Math. and Its Appl. E-ISSN: 2579-8936 P-ISSN: 1829-605X Vol. 15, No. 1, Maret 2018, 1-15.
- [25] Dellaert, Frank. (2002). "The Expectation Maximization Algorithm". Georgia Institute of Technology.
- [26] Wikipedia. *Cosine Similarity*, https://en.wikipedia.org/wiki/Cosine_similarity (diakses pada 15 April 2018).

LAMPIRAN A

Data Clustering Kata

TABEL A

Keanggotaan Kata pada Tiap Cluster (*Summarize with string*
dengan artikel tentang *chemical attack* di Douma)

Iterasi	Cluster 1	Cluster 2
Ke-0	['ultimately' 'ground' 'convulsions' 'killing' 'al' 'april' 'activist' 'deleted' 'helicopter' '2013' 'moscow' 'force' 'exposure' 'confirmed' 'details' 'works' 'outlets' 'france' 'failed' 'unwavering' 'syria' 'media' 'islam' 'mixed' 'mission' 'foaming' 'bomb' 'rural' 'independent' 'mouth' 'friday' 'sana' 'town' 'department' 'deaths' '000' 'dropped' 'allegedly' 'carried' 'symptoms' 'responsible' 'trying' 'verification' 'chemical' 'responsibility' 'nerve' 'raed' 'rebel' 'fabrication' 'opcw' 'held' 'saleh' 'us' 'put' 'consistent' 'suffered' 'blamed' 'reach' 'specialty' 'members' 'russia' 'said' 'everal' 'barrel' 'agent' 'region' 'talks' 'hospitals' '1' 'eastern' 'organization' 'negotiations' 'groups' 'bodies' 'let' 'shelling' 'assault' 'reports' 'obstruct' '2017' 'prohibition' 'rescuers' 'jaish' 'several' 'containing' 'ghouta' 'attack' 'opposition' 'attacks' 'siege' 'aerial' 'control' 'effects' 'exposed' 'seventy' 'contained' 'remain' 'suffocating' 'rockets' 'tweet' 'un' 'likely' 'powers' 'joint' 'volunteer' 'center' 'overnight' 'allegations' 'toxic' 'secure' 'people' 'launched' 'main' 'charity' 'rebels' 'hundreds' 'rise' 'making' 'sarin' 'broke' 'death' 'showing' 'last' 'spokeswoman' 'based' 'august' 'state' 'terrorists' 'pro' 'impossible' 'head' '75' 'forces'	['ms' '500' 'continuing' 'intervention' 'daughter' 'house' 'suggest' 'fired' 'discuss' 'smelled' 'may' '70' 'far' 'society' 'condemnation' 'deal' 'strike' 'following' 'inside' 'first' 'verified' 'television' 'abc' 'adding' 'launch' 'gross' 'investigations' 'immediate' 'also' 'julie' 'crisis' 'iran' 'responders' 'organizations' 'activity' 'price' 'anywhere' 'poses' 'affairs' 'referring' 'pressure' 'prepared' 'nothing' 'use' 'union' 'left' 'adviser' 'killed' 'medical' 'mindless' 'centre' 'world' 'turkey' 'helmets' 'including' 'base' 'west' 'accusing' 'chlorine' 'expected' 'images' 'forced' 'pretexts' 'came' 'widespread' 'substance' 'justify' 'response' 'form' 'fully' 'allies' 'global' 'treated' 'children' 'families' 'tool' 'russian' 'seen' 'abhorrent' 'action' 'sprawled' 'bashar' 'hospital' 'wake' 'matter' 'weapons' 'macron' 'want' 'identify' 'breathing' 'deter' 'bombings' 'ritual' 'military' 'situation' 'slam' 'evidence' 'men' 'hosed' 'emmanuel' 'armed' 'say' 'disturbing' 'hours' 'nations' 'examine' 'warning' 'possible' 'bishop' 'crying' 'hiding' 'upon'

	<p>'bears' 'agency' 'western' 'intense' 'arab' 'inquiry' 'tweeted' 'would' 'government' 'security' 'support' 'collapse' '150' 'advances' 'syrian' '80' 'monitoring' 'medics' 'died' 'invented' 'stay' 'douma']</p>	<p>'circumstances' 'another' 'rained' 'issued' 'hard' 'closest' 'extermination' 'year' 'twitter' 'statement' 'years' 'care' 'according' 'work' 'photos' 'makeshift' 'sparked' 'happened' 'concern' 'questions' 'donald' 'day' 'putin' 'assad' 'white' 'using' 'outrage' 'place' 'wrote' '24' 'fight' 'areas' 'dozens' 'parties' 'reason' 'retaliation' 'reported' 'violation' 'accept' 'table' 'saturday' 'asked' 'around' 'continue' 'footage' 'denied' 'shocking' 'airport' 'rescue' 'strikes' 'civilians' 'appared' 'graphic' 'defend' 'new' 'number' 'sources' 'centres' 'halt' 'accountability' 'american' 'french' 'backed' 'country' 'apply' 'vladimir' 'francis' 'workers' 'deny' 'fighters' 'cruise' 'women' 'poured' 'least' 'towards' 'father' 'bossert' 'united' 'incidents' 'organisation' 'stairwells' 'fabricated' 'homeland' 'whether' 'worth' 'council' 'independently' 'strongest' 'staff' 'among' 'took' 'horrific' 'bloodshed' 'mouths' 'clinics' 'way' 'establish' 'condemnations' 'revulsion' 'minister' 'law' 'emerged' 'airbase' 'air' 'beamed' 'shelters' 'basements' 'clearly' 'missile' 'reached' 'tom' 'yet' 'injured' 'international' 'known' 'warned' 'influence' 'inhumanity' 'prompted' 'conflict' 'eu' 'prevent' 'alleged' 'proliferation' 'clips' 'seeped' 'points' 'saying' 'trump' 'today' 'found' 'damascus' 'sunday' 'young' 'unverified' 'emergency' 'however' 'lifeless' 'fetched' 'difficulty' 'checks' 'actually' 'australia' 'recent' 'called' 'defence' 'struggle' 'accused' 'latest' 'foreign' 'assume' 'actual' 'backing' 'one' 'jean' 'pope' 'khan' 'suspected' 'buildings' 'coming' 'victims' 'much' 'terror' 'quickly'</p>
--	---	--

		<p>'populations' 'shows' 'believe' 'fabrications' 'fighting' 'enclave' 'involved' 'pay' 'attempt' 'video' 'gas' '85' 'subject' 'yves' 'speak' 'uk' 'evacuation' 'distressing' 'regime' 'bloc' 'simply' 'earlier' 'foam' 'lives' 'meanwhile' 'dead' 'bear' 'exactly' 'missiles' 'strongly' 'anything' 'escalation' 'grave' 'poisonous' 'evening' 'could' 'begging' 'used' 'anytime' 'responsibilities' 'related' 'lying' 'bbc' 'shortly' 'ally' 'relief' 'drian' 'verify' 'ministry' 'blockade' 'meet' 'big' 'maximum' 'suffocated' 'respond' 'test' 'many' 'civil' 'local' 'still' 'provocation' 'end' 'shape' 'apparent' 'commenting' 'condemned' 'terms' 'european' 'take' 'nostrils' 'almost' 'paris' 'continued' 'le' 'news' 'devices' 'resistance' 'animal' 'deadly' 'especially' 'survivors' 'mr' 'calls' 'sheikhoun' 'resolve' 'facilities' 'unit' 'unilaterally' 'mostly' 'told' 'deserves' 'humanitarian' 'late' 'army' 'runs' 'perpetrators' 'community' 'condemning' 'president' 'threat' 'monday' 'lashing' 'prolong' 'lashed' 'recorded' 'militant' 'homes' 'floors' 'condemns' 'defenceless']</p>
Ke-1	<p>[ultimately' 'intervention' 'fired' 'ground' 'c onvulsions' 'killing' 'april' 'activist' 'deal' 'first' 'deleted' 'helico pter' '2013' 'television' 'abc' 'adding' 'gross' 'moscow' ' responders' 'force' 'organizations' 'exposure' 'confirmed' 'refer ring' 'details' 'works' 'outlets' 'failed' 'adviser' 'unwavering' 'turk ey' 'islam' 'mixed' 'mission' 'pretexts' 'foaming' 'bomb' 'form' 'tool' 'rural' 'independent'</p>	<p>[ms' '500' 'continuing' 'daughter' 'h ouse' 'suggest' 'discuss' 'smelled' 'may' '70' 'al' 'far' 'society' 'condem nation' 'strike' 'following' 'inside' 'verified' 'launch' 'investiga tions' 'immediate' 'also' 'julie' 'crisis' 'iran' 'activity' 'price' 'anyw here' 'poses' 'affairs' 'pressure' 'prepared' 'nothing' 'use' 'union' 'fr ance' 'left' 'killed' 'medical' 'mindless' 'centre' 'syria' 'world' 'h elmets' 'media' 'including' 'base'</p>

<p>'russian' 'mouth' 'friday' 'sana' 'matter' 'dep artment' 'macron' 'deaths' '000' 'bombings' 'dropped' 'situation' 'eviden- ce' 'allegedly' 'carried' 'emmanuel' 'armed' 'say' 'symptoms' 'hours' 'nations' 'examine' 'responsible' 'trying' 'verification' 'closest' ' responsibility' 'nerve' 'raed' 'fabrication' 'opcw' 'years' 'according' 'work' 'sparked' 'concern' 'saleh' 'day' 'using' 'put' 'wrote' '24' 'fight' 'c- onsistent' 'areas' 'dozens' 'reason' 'retaliation' 'violation' 'suf- fered' 'blamed' 'accept' 'table' 'asked' 'denied' 'reach' 'specialty' 'm- embers' 'french' 'backed' 'vladimir' 'everal' 'barrel' 'agent' 'region' 'ta- lks' 'hospitals' '1' 'cruise' 'poured' 'least' 'towards' 'bossert' 'u- nited' 'incidents' 'eastern' 'fabricated' 'homeland' 'whether' ' organization' 'strongest' 'staff' 'negotiations' 'groups' 'bodies' 'way' ' establish' 'let' 'revulsion' 'shelling' 'assault' 'law' 'obstruct' '2017' 'airbase' 'prohibition' 'rescuers' 'jaish' 'several' 'cont- aining' 'tom' 'yet' 'ghouta' 'siege' 'influence' 'aerial' 'control' ' effects' 'eu' 'prevent' 'exposed' 'proliferation' 'seventy' 'points' 'c- ontained' 'remain' 'sunday' 'suffocating' 'rockets' 'tweet' 'fetched' 'un' 'l- ikely' 'powers' 'checks' 'joint' 'volunteer' 'center' 'overnight' 'allega- tions' 'toxic' 'secure' 'recent' 'launched' 'accused' 'main' 'charity' 'latest' 'rebels' 'assume' 'jean' 'hundreds' 'suspected' 'quickly' 'fabri- cations' 'rise' 'making' 'sarin' 'broke' 'death' 'showing' 'attempt' '8' 5' 'last' 'yves' 'spokeswoman' 'bloc' 'meanwhile' 'missile' s' 'anything' 'grave' 'poisonous' 'responsibilities' 'based' 'august' 'bbc' 'ally' 'state' 'ministry' 'meet' 'terrorists' 'still' 'provocation' 'pro' 'impossi- ble' 'shape' 'head' '75'</p>	<p>'west' 'accusing' 'chlorine' 'expecte- d' 'images' 'forced' 'came' 'widespread' 'substance' 'justify' 're- sponse' 'fully' 'allies' 'global' 'treated' 'children' 'families' 'seen' ' abhorrent' 'action' 'sprawled' 'bashar' 'hospital' 'wake' 'weapons' 'town' 'want' 'identify' 'breathing' 'deter' 'ritual' 'military' 'slam' 'men' 'hosed' 'disturbing' 'warning' 'possible' 'bishop' 'crying' 'hiding' ' upon' 'circumstances' 'chemical' 'another' 'rained' 'issued' 'hard' 'ext- ermination' 'year' 'twitter' 'statement' 'rebel' 'care' 'held' 'phot- os' 'makeshift' 'happened' 'questions' 'donald' 'putin' 'assad' ' us' 'white' 'outrage' 'place' 'parties' 'reported' 'saturday' 'aroun- d' 'continue' 'footage' 'shocking' 'airport' 'rescue' 'strikes' 'civilians' 'appeared' 'graphic' 'defend' 'new' 'number' 'sources' 'centres' 'h- alt' 'accountability' 'american' 'russia' 'country' 'apply' 'said' 'fran- cis' 'workers' 'deny' 'fighters' 'women' 'father' 'organisation' 'stai- rways' 'worth' 'council' 'independently' 'among' 'took' 'hor- rific' 'bloodshed' 'mouths' 'clinics' 'condemnations' 'minister' 'reports' 'emerged' 'air' 'beamed' 'shelters' 'basements' 'clearly' 'missile' 'reac- hed' 'injured' 'international' 'attack' 'known' 'opposition' 'attack- s' 'warned' 'inhumanity' 'prompted' 'conflict' 'alleged' 'clips' 'seeped' 's- aying' 'trump' 'today' 'found' 'damascus' 'young' 'unverified' 'em- ergency' 'however' 'lifeless' 'difficulty' 'actually' 'australia' 'peo- ple' 'called' 'defence' 'struggle' 'foreign' 'actual' 'backing' 'one' 'po- pe' 'khan' 'buildings' 'coming' 'victims' 'much' 'terror' 'population' s' 'shows' 'believe' 'fighting' 'enclave' 'involved' 'pay' 'video' 'ga- s' 'subject' 'speak' 'uk'</p>
---	--

	<p>'commenting' 'condemned' 'terms' 'forces' 'take' 'bears' 'agency' 'paris'</p> <p>'continued' 'news' 'western' 'intense' 'arab' 'inquiry' 'tweeted'</p> <p>'especially' 'calls' 'support' 'collapse' '150' 'unilaterally' 'advances'</p> <p>'humanitarian' 'late' 'army' 'perpetrators' 'threat' '80' 'monitoring'</p> <p>'medics' 'died' 'lashing' 'invented' 'stay' 'condemns']</p>	<p>'evacuation' 'distressing' 'regime' 's imply' 'earlier' 'foam' 'lives'</p> <p>'dead' 'bear' 'exactly' 'strongly' 'escalation' 'evening' 'could'</p> <p>'begging' 'used' 'anytime' 'related' 'lying' 'shortly' 'relief' 'drian'</p> <p>'verify' 'blockade' 'big' 'maximum' 'suffocated' 'respond' 'test' 'many'</p> <p>'civil' 'local' 'end' 'apparent' 'european' 'nostrils' 'almost' 'le'</p> <p>'devices' 'resistance' 'animal' 'deadly' 'survivors' 'mr' 'would'</p> <p>'government' 'sheikhoun' 'security' 'resolve' 'facilities' 'unit' 'mostly'</p> <p>'told' 'deserves' 'runs' 'community' 'syrian' 'condemning' 'president'</p> <p>'monday' 'prolong' 'lashed' 'recorded' 'militant' 'homes' 'floors' 'douna'</p> <p>'defenceless']</p>
Ke-2	<p>['ultimately' 'intervention' 'fired' 'ground' 'convulsions' 'killing' 'al'</p> <p>'april' 'activist' 'deal' 'first' 'deleted' 'helicopter' '2013'</p> <p>'television' 'abc' 'adding' 'gross' 'moscow' 'responders' 'force'</p> <p>'organizations' 'exposure' 'confirmed' 'referring' 'details' 'works'</p> <p>'outlets' 'failed' 'adviser' 'unwavering' 'syria' 'turkey' 'media' 'islam'</p> <p>'mixed' 'mission' 'pretexts' 'foaming' 'bomb' 'form' 'tool' 'rural'</p> <p>'independent' 'russian' 'mouth' 'friday' 'sanctions' 'matter' 'town'</p> <p>'department' 'macron' 'deaths' '000' 'bombings' 'dropped' 'situation'</p> <p>'evidence' 'allegedly' 'carried' 'emmanuel' 'armed' 'say' 'symptoms'</p> <p>'hours' 'nations' 'examine' 'responsible' 'trying' 'verification'</p> <p>'closest' 'responsibility' 'nerve' 'raed' 'rebel' 'fabrication' 'opcw'</p> <p>'years' 'held' 'according' 'work' 'sparked' 'concern' 'saleh' 'day'</p> <p>'using' 'put' 'wrote' '24' 'fight' 'consistent' 'areas' 'dozens' 'reason'</p> <p>'retaliation' 'violation' 'suffered' 'blamed' 'accept' 'table' 'asked'</p>	<p>['ms' '500' 'continuing' 'daughter' 'house' 'suggest' 'discuss' 'smelled'</p> <p>'may' '70' 'far' 'society' 'condemnation' 'strike' 'following' 'inside'</p> <p>'verified' 'launch' 'investigations' 'immediate' 'also' 'julie' 'crisis'</p> <p>'iran' 'activity' 'price' 'anywhere' 'poses' 'affairs' 'pressure'</p> <p>'prepared' 'nothing' 'use' 'union' 'france' 'left' 'killed' 'medical'</p> <p>'mindless' 'centre' 'world' 'helmets' 'including' 'base' 'west' 'accusing'</p> <p>'chlorine' 'expected' 'images' 'forced' 'came' 'widespread' 'substance'</p> <p>'justify' 'response' 'fully' 'allies' 'global' 'treated' 'children'</p> <p>'families' 'seen' 'abhorrent' 'action' 'sprawled' 'bashar' 'hospital'</p> <p>'wake' 'weapons' 'want' 'identify' 'breathing' 'deter' 'ritual' 'military'</p> <p>'slam' 'men' 'hosed' 'disturbing' 'warning' 'possible' 'bishop' 'crying'</p> <p>'hiding' 'upon' 'circumstances' 'chemical' 'another' 'rained' 'issued'</p> <p>'hard' 'extermination' 'year' 'twitter' 'statement' 'care' 'photos'</p> <p>'makeshift' 'happened' 'questions' 'donald' 'putin' 'assad' 'us' 'white'</p>

<p>'denied' 'reach' 'specialty' 'members' 'french' 'backed' 'said' 'vladimir' 'everal' 'barrel' 'agent' 'region' 'talks' 'hospitals' 'l' 'cruise' 'poured' 'least' 'towards' 'bossert' 'united' 'incidents' 'eastern' 'fabricated' 'homeland' 'whether' 'organization' 'strongest' 'staff' 'negotiations' 'groups' 'bodies' 'way' 'establish' 'let' 'revulsion' 'shelling' 'assault' 'law' 'obstruct' '2017' 'air base' 'prohibition' 'rescuers' 'jaish' 'several' 'containing' 'tom' 'yet' 'ghouta' 'opposition' 'attacks' 'siege' 'influence' 'aerial' 'control' 'effects' 'eu' 'prevent' 'exposed' 'proliferation' 'seventy' 'points' 'contained' 'remain' 'sunday' 'suffocating' 'rockets' 'tweet' 'fetched' 'un' 'likely' 'powers' 'checks' 'joint' 'volunteer' 'center' 'overnight' 'allegations' 'toxic' 'secure' 'people' 'recent' 'launched' 'accused' 'main' 'charity' 'latest' 'rebels' 'assume' 'jean' 'hundreds' 'khan' 'suspected' 'quickly' 'fabrications' 'rise' 'making' 'sarin' 'broke' 'death' 'showing' 'attempt' '85' 'last' 'yves' 'spokeswoman' 'bloc' 'meanwhile' 'missiles' 'anything' 'grave' 'poisonous' 'responsibilities' 'based' 'august' 'bbc' 'ally' 'state' 'ministry' 'meet' 'terrorists' 'still' 'provocation' 'pro' 'impossible' 'shape' 'head' '75' 'commenting' 'condemned' 'terms' 'forces' 'take' 'bears' 'agency' 'paris' 'continued' 'news' 'western' 'intense' 'arab' 'inquiry' 'tweeted' 'especially' 'government' 'calls' 'sheikhoun' 'support' 'collapse' '150' 'unilaterally' 'told' 'advances' 'humanitarian' 'late' 'army' 'perpetrators' 'threat' '80' 'monitoring' 'medics' 'died' 'lashing' 'invented' 'stay' 'condemns' 'duma']</p>	<p>'outrage' 'place' 'parties' 'reported' 'saturday' 'around' 'continue' 'footage' 'shocking' 'airport' 'rescue' 'strikes' 'civilians' 'appeared' 'graphic' 'defend' 'new' 'number' 'sources' 'centres' 'halt' 'accountability' 'american' 'russia' 'country' 'apply' 'francis' 'workers' 'deny' 'fighters' 'women' 'father' 'organization' 'stairwells' 'worth' 'council' 'independently' 'among' 'took' 'horrific' 'bloodshed' 'mouths' 'clinics' 'condemnations' 'minister' 'reports' 'emerged' 'air' 'beamed' 'shelters' 'basements' 'clearly' 'missile' 'reached' 'injured' 'international' 'attack' 'known' 'warned' 'inhumanity' 'prompted' 'conflict' 'alleged' 'clips' 'seeped' 'saying' 'trump' 'today' 'found' 'damascus' 'young' 'unverified' 'emergency' 'however' 'lifeless' 'difficulty' 'actually' 'australia' 'called' 'defence' 'struggle' 'foreign' 'actual' 'backing' 'one' 'people' 'buildings' 'coming' 'victims' 'much' 'terror' 'populations' 'shows' 'believe' 'fighting' 'enclave' 'involved' 'pay' 'video' 'gas' 'subject' 'speak' 'uk' 'evacuation' 'distressing' 'regime' 'simply' 'earlier' 'foam' 'lives' 'dead' 'bear' 'exactly' 'strongly' 'escalation' 'evening' 'could' 'begging' 'used' 'anytime' 'related' 'lying' 'shortly' 'relief' 'drian' 'verify' 'blockade' 'big' 'maximum' 'suffocated' 'respond' 'test' 'many' 'civil' 'local' 'end' 'apparent' 'european' 'nostrils' 'almost' 'le' 'devices' 'resistance' 'animal' 'deadly' 'survivors' 'mr' 'would' 'security' 'resolve' 'facilities' 'unit' 'mostly' 'deserves' 'runs' 'community' 'syrian' 'condemning' 'president' 'monday' 'prolong' 'lashed' 'recorded' 'militant' 'homes' 'floors' 'defenceless']</p>
---	--

Ke-3	<p> [ultimately' 'fired' 'ground' 'convulsions' 'killing' '70' 'al' 'april' 'activist' 'deal' 'deleted' 'helicopter' '2013' 'moscow' 'force' 'organizations' 'exposure' 'confirmed' 'details' 'works' 'outlets' 'failed' 'unwavering' 'syria' 'helmets' 'media' 'islam' 'mixed' 'mission' 'foaming' 'bomb' 'rural' 'independent' 'russian' 'mouth' 'friday' 'sana' 'town' 'department' 'deaths' '000' 'dropped' 'allegedly' 'carried' 'say' 'symptoms' 'responsible' 'trying' 'verification' 'responsibility' 'nerve' 'raed' 'rebel' 'fabrication' 'opcw' 'held' 'saleh' 'us' 'put' 'consistent' 'areas' 'dozens' 'suffered' 'blamed' 'asked' 'each' 'specialty' 'members' 'number' 'backed' 'said' 'everal' 'barrel' 'age' 'region' 'talks' 'hospitals' '1' 'least' 'eastern' 'organization' 'negotiations' 'groups' 'bodies' 'let' 'shelling' 'assault' 'reports' 'obstruct' '2017' 'prohibition' 'rescuers' 'jaish' 'several' 'containing' 'ghouta' 'opposition' 'attacks' 'siege' 'aerial' 'control' 'effects' 'alleged' 'exposed' 'seventy' 'contained' 'remain' 'sunday' 'suffocating' 'rockets' 'tweet' 'un' 'likely' 'powers' 'joint' 'volunteer' 'center' 'overnight' 'allegations' 'toxic' 'secure' 'people' 'launched' 'main' 'charity' 'rebels' 'hundreds' 'khan' 'suspected' 'fabrications' 'rise' 'making' 'sarin' 'broke' 'death' 'showing' 'attempt' 'last' 'spokeswoman' 'based' 'august' 'bbc' 'ally' 'state' 'terrorists' 'still' 'por' 'impossible' 'head' '75' 'forces' 'bears' 'agency' 'continued' 'news' 'western' 'intense' 'arab' 'inquiry' 'tweeted' 'would' 'government' 'sheikhoun' 'security' 'support' 'collapse' '150' 'told' 'advances' 'army' '80' 'monitoring' 'medics' 'died' 'invented' 'stay' 'douma'] </p>	<p> [ms' '500' 'continuing' 'intervention' 'daughter' 'house' 'suggest' 'discuss' 'smelled' 'may' 'far' 'society' 'condemnation' 'strike' 'following' 'inside' 'first' 'verified' 'television' 'abc' 'adding' 'launch' 'gross' 'investigations' 'immediate' 'also' 'julie' 'crisis' 'iran' 'responders' 'activity' 'price' 'anywhere' 'poses' 'affairs' 'referring' 'pressure' 'prepared' 'nothing' 'use' 'union' 'france' 'left' 'adviser' 'killed' 'medical' 'minds' 'centre' 'world' 'turkey' 'including' 'base' 'west' 'accusing' 'chlorine' 'expected' 'images' 'forced' 'pretexts' 'came' 'widespread' 'substance' 'justify' 'response' 'form' 'fully' 'allies' 'global' 'treated' 'children' 'families' 'tool' 'seen' 'abhorrent' 'action' 'sprawled' 'basha' 'hospital' 'wake' 'matter' 'weapons' 'macron' 'want' 'identify' 'breathing' 'deter' 'bombings' 'ritual' 'military' 'situation' 'slam' 'evidence' 'men' 'hosed' 'emmanuel' 'armed' 'disturbing' 'hours' 'nations' 'examine' 'warning' 'possible' 'bishop' 'crying' 'hiding' 'upon' 'circumstances' 'chemical' 'another' 'rained' 'issued' 'hard' 'closest' 'extinction' 'year' 'twitter' 'statement' 'years' 'care' 'according' 'work' 'photos' 'makeshift' 'sparked' 'happened' 'concern' 'questions' 'donald' 'day' 'putin' 'assad' 'white' 'using' 'outrage' 'place' 'wrote' '24' 'fight' 'parties' 'reason' 'retaliation' 'reported' 'violation' 'accept' 'table' 'saturday' 'around' 'continue' 'footage' 'denied' 'shocking' 'airport' 'rescue' 'strikes' 'civilians' 'appeared' 'graphic' 'defend' 'new' 'sources' 'centres' 'halt' 'accountability' 'american' 'russia' 'french' 'country' 'apply' 'vladimir' 'francis' 'workers' 'deny' 'fighters' 'cruise' 'women' 'poured' 'towards' </p>
------	--	---

		<p>'father' 'bossert' 'united' 'incidents' 'organisation' 'stairwells' 'fabricated' 'homeland' 'whether' 'worth' 'council' 'independently' 'strongest' 'staff' 'among' 'took' 'horrific' 'bloodshed' 'mouths' 'clinics' 'way' 'establish' 'condemnations' 'revulsion' 'minister' 'law' 'emerged' 'airbase' 'air' 'beamed' 'shelters' 'basements' 'clearly' 'missile' 'reached' 'tom' 'yet' 'injured' 'international' 'attack' 'known' 'warned' 'influence' 'inhumanity' 'prompted' 'conflict' 'eu' 'prevent' 'proliferation' 'clips' 'seeped' 'points' 'saying' 'trump' 'today' 'found' 'damascus' 'young' 'unverified' 'emergency' 'however' 'lifeless' 'fetched' 'difficulty' 'checks' 'actually' 'australia' 'recent' 'called' 'defence' 'struggle' 'accused' 'latest' 'foreign' 'assume' 'actual' 'backing' 'one' 'jean' 'pope' 'buildings' 'coming' 'victims' 'much' 'terror' 'quickly' 'populations' 'shows' 'believe' 'fighting' 'enclave' 'involved' 'pay' 'video' 'gas' '85' 'subject' 'yves' 'speak' 'uk' 'evacuation' 'distressing' 'regime' 'bloc' 'simply' 'earlier' 'foam' 'lives' 'meanwhile' 'dead' 'bear' 'exactly' 'missiles' 'strongly' 'anything' 'escalation' 'grave' 'poisonous' 'evening' 'could' 'begging' 'used' 'anytime' 'responsibilities' 'related' 'lying' 'shortly' 'relief' 'drian' 'verify' 'ministry' 'blockade' 'meet' 'big' 'maximum' 'suffocated' 'respond' 'test' 'many' 'civil' 'local' 'provocation' 'end' 'shape' 'apparent' 'commenting' 'condemned' 'terms' 'european' 'take' 'nostrils' 'almost' 'paris' 'le' 'devices' 'resistance' 'animal' 'deadly' 'especially' 'survivors' 'mr' 'calls' 'resolve' 'facilities' 'unit' 'unilaterally' 'mostly' 'deserves' 'humanitarian' 'late' 'runs'</p>
--	--	--

		'perpetrators' 'community' 'syrian' 'condemning' 'president' 'threat' 'monday' 'lashing' 'prolong' 'lashed' 'recorded' 'militant' 'homes' 'floors' 'condemns' 'defenceless']
Ke-4	[ultimately' 'fired' 'ground' 'convulsions' 'killing' '70' 'al' 'april' 'activist' 'deal' 'deleted' 'helicopter' '2013' 'moscow' 'force' 'organizations' 'exposure' 'confirmed' 'details' 'works' 'outlets' 'union' 'failed' 'unwavering' 'syria' 'helmets' 'media' 'islam' 'chlorine' 'images' 'mixed' 'mission' 'foaming' 'bomb' 'rural' 'independent' 'russian' 'mouth' 'friday' 'sana' 'hospital' 'town' 'department' 'deaths' '000' 'dropped' 'allegedly' 'carried' 'say' 'symptoms' 'responsible' 'trying' 'verification' 'responsibility' 'nerve' 'raed' 'rebel' 'fabrication' 'opcw' 'held' 'saleh' 'us' 'put' 'consistent' 'areas' 'dozens' 'suffered' 'blamed' 'asked' 'reach' 'specialty' 'graphic' 'members' 'number' 'backed' 'said' 'everal' 'barrel' 'agent' 'region' 'talks' 'hospitals' 'i' 'least' 'eastern' 'organization' 'negotiations' 'groups' 'bodies' 'let' 'shelling' 'assault' 'reports' 'obstruct' '2017' 'emerged' 'prohibition' 'rescuers' 'jaish' 'basements' 'several' 'containing' 'ghouta' 'opposition' 'attacks' 'siege' 'aerial' 'control' 'effects' 'alleged' 'exposed' 'seventy' 'contained' 'remain' 'damascus' 'sunday' 'suffocating' 'rockets' 'tweet' 'un' 'likely' 'powers' 'joint' 'volunteer' 'center' 'overnight' 'allegations' 'toxic' 'secure' 'people' 'launched' 'main' 'charity' 'rebels' 'hundreds' 'khan' 'suspected' 'victims' 'fabrications' 'rise' 'making' 'sarin' 'broke' 'death' 'showing' 'attempt' 'last' 'spokeswoman' 'earlier' 'based' 'august' 'bbc' 'ally' 'state' 'suffocated' 'terrorists' 'local' 'still' 'proof' 'impossible'	['ms' '500' 'continuing' 'intervention' 'daughter' 'house' 'suggest' 'discuss' 'smelled' 'may' 'far' 'society' 'condemnation' 'strike' 'following' 'inside' 'first' 'verified' 'television' 'abc' 'adding' 'launch' 'gross' 'investigations' 'immediate' 'also' 'julie' 'crisis' 'iran' 'responders' 'activity' 'price' 'anywhere' 'poses' 'affairs' 'referring' 'pressure' 'prepared' 'nothing' 'use' 'france' 'left' 'adviser' 'killed' 'medical' 'mindless' 'centre' 'world' 'turkey' 'including' 'base' 'west' 'accusing' 'expected' 'forced' 'pretexts' 'came' 'widespread' 'substance' 'justify' 'response' 'form' 'fully' 'allies' 'global' 'treated' 'children' 'families' 'tool' 'seen' 'abhorrent' 'action' 'sprawled' 'bashar' 'wake' 'matter' 'weapons' 'macron' 'want' 'identify' 'breathing' 'deter' 'bombings' 'ritual' 'military' 'situation' 'slam' 'evidence' 'men' 'hosed' 'emmanuel' 'armed' 'disturbing' 'hours' 'nations' 'examine' 'warning' 'possible' 'bishop' 'crying' 'hiding' 'upon' 'circumstances' 'chemical' 'another' 'rained' 'issued' 'hard' 'closest' 'extermination' 'year' 'twitter' 'statement' 'years' 'care' 'according' 'work' 'photos' 'makeshift' 'sparked' 'happened' 'concern' 'questions' 'donald' 'day' 'putin' 'assad' 'white' 'using' 'outrage' 'place' 'wrote' '24' 'fight' 'parties' 'reason' 'retaliation' 'reported' 'violation' 'accept' 'table' 'saturday' 'around' 'continue' 'footage' 'denied' 'shocking' 'airport' 'rescue' 'strikes' 'civilians' 'appeared' 'defend' 'new' 'sources' 'centres' 'halt' 'accountability' 'american' 'russia'

	<p>'head' '75' 'forces' 'bears' 'agency' 'continue d' 'news' 'western'</p> <p>'intense' 'arab' 'inquiry' 'tweeted' 'would' 'g overnment' 'sheikhoun'</p> <p>'security' 'support' 'collapse' '150' 'told' 'ad vances' 'army' '80'</p> <p>'monitoring' 'medics' 'died' 'invented' 'stay' 'douma']</p>	<p>'french' 'country' 'apply' 'vladimir' 'francis' 'workers' 'deny'</p> <p>'fighters' 'cruise' 'women' 'poured' ' towards' 'father' 'bossert' 'united'</p> <p>'incidents' 'organisation' 'stairwells' ' 'fabricated' 'homeland' 'whether'</p> <p>'worth' 'council' 'independently' 'st rongest' 'staff' 'among' 'took'</p> <p>'horrific' 'bloodshed' 'mouths' 'clin ics' 'way' 'establish'</p> <p>'condemnations' 'revulsion' 'minist er' 'law' 'airbase' 'air' 'beamed'</p> <p>'shelters' 'clearly' 'missile' 'reached' ' 'tom' 'yet' 'injured'</p> <p>'international' 'attack' 'known' 'war ned' 'influence' 'inhumanity'</p> <p>'prompted' 'conflict' 'eu' 'prevent' ' proliferation' 'clips' 'seeped'</p> <p>'points' 'saying' 'trump' 'today' 'fou nd' 'young' 'unverified' 'emergency'</p> <p>'however' 'lifeless' 'fetched' 'diffic ulty' 'checks' 'actually'</p> <p>'australia' 'recent' 'called' 'defence' 'struggle' 'accused' 'latest'</p> <p>'foreign' 'assume' 'actual' 'backing' 'one' 'jean' 'pope' 'buildings'</p> <p>'coming' 'much' 'terror' 'quickly' 'p opulations' 'shows' 'believe'</p> <p>'fighting' 'enclave' 'involved' 'pay' ' video' 'gas' '85' 'subject' 'yves'</p> <p>'speak' 'uk' 'evacuation' 'distressin g' 'regime' 'bloc' 'simply' 'foam'</p> <p>'lives' 'meanwhile' 'dead' 'bear' 'ex actly' 'missiles' 'strongly'</p> <p>'anything' 'escalation' 'grave' 'pois onous' 'evening' 'could' 'begging'</p> <p>'used' 'anytime' 'responsibilities' 'r elated' 'lying' 'shortly' 'relief'</p> <p>'drian' 'verify' 'ministry' 'blockade' 'meet' 'big' 'maximum' 'respond'</p> <p>'test' 'many' 'civil' 'provocation' 'en d' 'shape' 'apparent' 'commenting'</p> <p>'condemned' 'terms' 'european' 'tak e' 'nostrils' 'almost' 'paris' 'le'</p> <p>'devices' 'resistance' 'animal' 'dead ly' 'especially' 'survivors' 'mr'</p> <p>'calls' 'resolve' 'facilities' 'unit' 'uni laterally' 'mostly' 'deserves'</p>
--	---	--

		'humanitarian' 'late' 'runs' 'perpetrators' 'community' 'syrian' 'condemning' 'president' 'threat' 'monday' 'lashing' 'prolong' 'lashed' 'recorded' 'militant' 'homes' 'floors' 'condemns' 'defenceless']
Ke-5	<p>[ultimately' 'fired' 'ground' 'convulsions' 'killing' '70' 'al' 'april'</p> <p>'activist' 'deal' 'deleted' 'helicopter' '2013' 'moscow' 'force'</p> <p>'organizations' 'exposure' 'confirmed' 'details' 'works' 'outlets' 'union'</p> <p>'failed' 'unwavering' 'syria' 'helmets' 'media' 'islam' 'chlorine'</p> <p>'images' 'mixed' 'mission' 'foaming' 'bomb' 'rural' 'independent'</p> <p>'russian' 'mouth' 'friday' 'sana' 'hospital' 'town' 'department' 'deaths'</p> <p>'000' 'dropped' 'allegedly' 'carried' 'say' 'symptoms' 'responsible'</p> <p>'trying' 'verification' 'responsibility' 'nerve' 'raed' 'rebel'</p> <p>'fabrication' 'opcw' 'held' 'saleh' 'us' 'put' 'consistent' 'areas'</p> <p>'dozens' 'suffered' 'blamed' 'asked' 'reach' 'specialty' 'graphic'</p> <p>'members' 'number' 'backed' 'said' 'everal' 'barrel' 'agent' 'region'</p> <p>'talks' 'hospitals' 'l' 'least' 'eastern' 'organization' 'negotiations'</p> <p>'groups' 'bodies' 'let' 'shelling' 'assault' 'reports' 'obstruct' '2017'</p> <p>'emerged' 'prohibition' 'rescuers' 'jaish' 'basements' 'several'</p> <p>'containing' 'ghouta' 'opposition' 'attacks' 'siege' 'aerial' 'control'</p> <p>'effects' 'alleged' 'exposed' 'seventy' 'contained' 'remain' 'damascus'</p> <p>'sunday' 'suffocating' 'rockets' 'tweet' 'un' 'likely' 'powers' 'joint'</p> <p>'volunteer' 'center' 'overnight' 'allegations' 'toxic' 'secure' 'people'</p> <p>'launched' 'main' 'charity' 'rebels' 'hundreds' 'khan' 'suspected'</p> <p>'victims' 'fabrications' 'rise' 'making' 'sarin' 'broke' 'death' 'showing'</p> <p>'attempt' 'last' 'spokeswoman' 'earlier' 'based' 'august' 'bbc' 'ally'</p>	<p>['ms' '500' 'continuing' 'intervention' 'daughter' 'house' 'suggest'</p> <p>'discuss' 'smelled' 'may' 'far' 'society' 'condemnation' 'strike'</p> <p>'following' 'inside' 'first' 'verified' 'television' 'abc' 'adding'</p> <p>'launch' 'gross' 'investigations' 'immediate' 'also' 'julie' 'crisis'</p> <p>'iran' 'responders' 'activity' 'price' 'anywhere' 'poses' 'affairs'</p> <p>'referring' 'pressure' 'prepared' 'nothing' 'use' 'france' 'left'</p> <p>'adviser' 'killed' 'medical' 'mindless' 'centre' 'world' 'turkey'</p> <p>'including' 'base' 'west' 'accusing' 'expected' 'forced' 'pretexts' 'came'</p> <p>'widespread' 'substance' 'justify' 'response' 'form' 'fully' 'allies'</p> <p>'global' 'treated' 'children' 'families' 'tool' 'seen' 'abhorrent'</p> <p>'action' 'sprawled' 'bashar' 'wake' 'matter' 'weapons' 'macron' 'want'</p> <p>'identify' 'breathing' 'deter' 'bombings' 'ritual' 'military' 'situation'</p> <p>'slam' 'evidence' 'men' 'hosed' 'emmanuel' 'armed' 'disturbing' 'hours'</p> <p>'nations' 'examine' 'warning' 'possible' 'bishop' 'crying' 'hiding' 'upon'</p> <p>'circumstances' 'chemical' 'another' 'rained' 'issued' 'hard' 'closest'</p> <p>'extermination' 'year' 'twitter' 'statement' 'years' 'care' 'according'</p> <p>'work' 'photos' 'makeshift' 'sparked' 'happened' 'concern' 'questions'</p> <p>'donald' 'day' 'putin' 'assad' 'white' 'using' 'outrage' 'place' 'wrote'</p> <p>'24' 'fight' 'parties' 'reason' 'retaliation' 'reported' 'violation'</p> <p>'accept' 'table' 'saturday' 'around' 'continue' 'footage' 'denied'</p> <p>'shocking' 'airport' 'rescue' 'strikes' 'civilians' 'appeared' 'defend'</p>

	<p>'state' 'suffocated' 'terrorists' 'local' 'still' 'p ro' 'impossible' 'head' '75' 'forces' 'bears' 'agency' 'continue d' 'news' 'western' 'intense' 'arab' 'inquiry' 'tweeted' 'would' 'g overnment' 'sheikhoun' 'security' 'support' 'collapse' '150' 'told' 'ad vances' 'army' '80' 'monitoring' 'medics' 'died' 'invented' 'stay' 'douma']</p>	<p>'new' 'sources' 'centres' 'halt' 'acco untability' 'american' 'russia' 'french' 'country' 'apply' 'vladimir' 'francis' 'workers' 'deny' 'fighters' 'cruise' 'women' 'poured' ' towards' 'father' 'bossert' 'united' 'incidents' 'organisation' 'stairwells' ' 'fabricated' 'homeland' 'whether' 'worth' 'council' 'independently' 'st rongest' 'staff' 'among' 'took' 'horrific' 'bloodshed' 'mouths' 'clin ics' 'way' 'establish' 'condemnations' 'revulsion' 'minist er' 'law' 'airbase' 'air' 'beamed' 'shelters' 'clearly' 'missile' 'reached' ' 'tom' 'yet' 'injured' 'international' 'attack' 'known' 'war ned' 'influence' 'inhumanity' 'prompted' 'conflict' 'eu' 'prevent' ' proliferation' 'clips' 'seeped' 'points' 'saying' 'trump' 'today' 'fou nd' 'young' 'unverified' 'emergency' 'however' 'lifeless' 'fetched' 'diffic ulty' 'checks' 'actually' 'australia' 'recent' 'called' 'defence' 'struggle' 'accused' 'latest' 'foreign' 'assume' 'actual' 'backing' 'one' 'jean' 'pope' 'buildings' 'coming' 'much' 'terror' 'quickly' 'p opulations' 'shows' 'believe' 'fighting' 'enclave' 'involved' 'pay' ' video' 'gas' '85' 'subject' 'yves' 'speak' 'uk' 'evacuation' 'distressin g' 'regime' 'bloc' 'simply' 'foam' 'lives' 'meanwhile' 'dead' 'bear' 'ex actly' 'missiles' 'strongly' 'anything' 'escalation' 'grave' 'pois onous' 'evening' 'could' 'begging' 'used' 'anytime' 'responsibilities' 'r elated' 'lying' 'shortly' 'relief' 'drian' 'verify' 'ministry' 'blockade' 'meet' 'big' 'maximum' 'respond' 'test' 'many' 'civil' 'provocation' 'en d' 'shape' 'apparent' 'commenting' 'condemned' 'terms' 'european' 'tak e' 'nostrils' 'almost' 'paris' 'le' 'devices' 'resistance' 'animal' 'dead ly' 'especially' 'survivors' 'mr'</p>
--	--	--

		'calls' 'resolve' 'facilities' 'unit' 'uni laterally' 'mostly' 'deserves' 'humanitarian' 'late' 'runs' 'perpetra tors' 'community' 'syrian' 'condemning' 'president' 'threat' 'm onday' 'lashing' 'prolong' 'lashed' 'recorded' 'militant' 'homes' 'floors' 'condemns' 'defenceless']
--	--	--

TABEL B

Iterasi	Titik Tengah Cluster 1 (a)	Titik Tengah Cluster 2 (b)
Ke-0	[0.000000 0.000000 0.200000 0.000000]	[0.200000 0.000000 0.000000 0.200000]
Ke-1	[0.00152657 0.00124047 0.00095616 0.00256968]	[0.00266347 0.00313531 0.00400454 0.0018482]
Ke-2	[0.00153031 0.00139214 0.00132006 0.00199724]	[0.00164544 0.00177988 0.00206866 0.00140417]
Ke-3	[0.0015578 0.00152297 0.0015614 0.0017852]	[0.00158856 0.00161683 0.00177058 0.00155057]
Ke-4	[0.00156721 0.00155774 0.00163411 0.00170799]	[0.00157808 0.0015807 0.00169403 0.00162169]
Ke-5	[0.00157137 0.00156728 0.00165613 0.00168082]	[0.00157576 0.00157298 0.00167373 0.00165017]
Ke-6	[0.00157317 0.00156992 0.00166283 0.00167134]	[0.001575 0.00157137 0.00166811 0.00166067]
Ke-7	[0.00157392 0.00157067 0.00166488 0.00166806]	[0.00157467 0.00157105 0.0016665 0.00166439]
Ke-8	[0.00157423 0.00157088 0.00166552 0.00166694]	[0.00157452 0.00157099 0.00166603 0.00166568]
Ke-9	[0.00157434 0.00157095 0.00166572 0.00166655]	[0.00157446 0.00157098 0.00166589 0.00166612]

TABEL C

Iterasi	Cluster 1	Cluster 2
Ke-0	<p>['ms' 'next' 'journalist' 'sentenced' 'prosecutors' 'murder' 'kim' 'media' '30' 'madsen' 'homemade' 'discovered' 'previously' 'school' 'year' 'journalism' 'october' 'day' 'board' 'said' 'police' 'described' 'vessel' 'sweden' 'remains' 'gave' 'mutilated' 'copyright' 'dismembering' 'body' 'know' 'last' 'submarine' '21' 'dismembered' 'august' 'went' 'self' 'night' 'inventor' 'caption' 'wall' 'built' 'image']</p>	<p>['40' '6' 'tortured' 'coast' 'motivated' 'plastic' 'formidable' 'may' 'driven' 'deserve' 'burkoe' 'far' 'offer' 'argued' 'london' 'following' 'inside' 'unsupported' 'aboard' 'first' 'court' 'equates' 'empathy' 'story' 'return' 'freelance' 'city' '1987' 'emotionally' 'hald' 'outside' 'also' 'anna' 'jepsen' 'taken' 'ended' 'crimes' 'given' 'studied' 'oxbridge' 'danish' 'south' 'finally' 'activity' 'legs' 'raised' 'overboard' 'necessary' 'divide' 'films' 'born' 'removed' 'details' 'pressure' 'indecent' 'koge' 'ordered' 'help' 'undocumented' 'missing' 'crime' 'appear' 'personal' 'killed' 'aid' '50km' 'defense' 'including' 'danger' 'clear' 'carbon' 'exhibited' 'images' 'semen' 'ap' 'accident' 'corpse' 'case' 'jurors' 'hit' 'sought' 'woman' 'horrible' 'various' '47' 'relations' 'manner' 'seen' 'guilt' 'written' 'friends' 'metal' 'team' 'weighted' 'offering' 'committed' 'read' 'town' 'extended' 'press' 'never' 'private' 'rado' 'deaths' 'pleaded' 'peter' 'bags' 'honours' 'trelleborg' 'motive' 'dropped' 'throat' 'planned' 'character' 'tendencies' 'quoted' 'severe' 'evidence' 'uc3' 'via' 'say' 'baltic' 'hours' 'legal' 'make' '10' 'early' 'conviction' '12' 'piece' 'sea' 'accepted' '8' 'respect' 'knit' 'hatch' 'wanted' 'claim' 'beijing' 'bury' 'interview' 'torture' 'added' 'established' 'venture' 'explanations' 'within' 'multiple' 'sentence' 'explanation' 'fell' 'boyfriend' 'credible' 'dividing' 'terrible' 'years']</p>

		<p> 'parts' 'wonderful' 'according' 'photos' 'detailed' 'happened' 'genitals' 'masters' 'arrest' 'bay' 'magazine' 'prior' 'betina' 'premeditated' 'place' 'problem' 'set' 'lawyer' 'excerpts' 'consistent' 'telling' 'classmate' 'something' 'cyclist' 'moved' 'top' 'close' 'economics' 'account' 'events' 'opening' 'quietly' 'denied' 'verdict' 'university' 'handling' 'interviewed' 'suffocate' 'unconvincing' 'burying' 'researching' 'new' 'publications' 'bax' 'later' '16' 'sexual' 'arguing' 'randomly' 'report' 'develops' 'remorse' 'showed' 'departed' 'spacecraft' 'traits' 'region' 'suspicion' 'bed' 'women' 'codrea' 'decision' 'admitted' 'accidental' 'brutal' 'featured' 'either' 'family' 'lindhardt' 'sexually' 'decided' 'getty' 'playback' 'go' 'whether' 'trial' 'explaining' 'island' 'along' 'throwing' 'took' 'spare' 'timeline' 'somewhere' 'monoxide' 'judge' 'clothing' '11' 'winning' 'swedish' 'assault' 'stabbed' 'psychopathic' 'device' 'computer' 'parole' 'determined' '2017' 'acted' 'emotion' 'conceal' 'differing' 'agreed' 'confessed' 'nordic' 'changed' 'prosecutor' 'dismember' 'even' 'afp' 'month' 'several' 'reached' 'sank' 'international' 'caused' 'stopped' 'session' 'psychiatric' 'programme' 'aug' 'underwear' 'waters' 'testifying' 'authorities' 'ride' 'tried' 'deliberately' 'saying' 'spotted' 'kill' 'found' 'convict' 'suffocating' 'horror' 'restrictions' 'offered' 'charged' 'buch' 'lifeless' 'march' 'impaired' 'anger' 'conducted' 'helped' 'good' 'accounts' 'made' 'impaled' </p>
--	--	---

		'denmark' 'arms' 'upwards' 'sailing' 'trustworthy' 'lack' 'sky' 'cause' 'pitch' 'allegations' 'toxic' 'investigators' 'manned' 'sudden' 'pieces' 'accused' 'old' 'videos' 'open' 'columbia' 'associated' 'nautilus' 'one' 'boarding' 'since' 'abroad' 'co' 'immediately' 'walls' 'dumped' 'southern' 'shows' 'believe' 'heard' 'accidentally' 'disappeared' 'anette' 'prison' 'class' 'york' 'gruesome' 'across' 'death' 'showing' '22' 'special' 'life' '150lb' 'smaller' 'hearing' 'shifting' 'wounds' 'guardian' 'cohort' 'departing' 'chosen' 'large' 'lift' 'strait' 'deck' 'tools' 'give' 'grew' 'reuters' 'dead' 'risk' 'tied' 'stockings' 'guilty' 'evening' 'could' 'sink' 'jakob' 'times' 'passing' 'based' 'opinion' 'founder' 'bbc' 'friend' 'weeks' 'brought' 'big' 'tonne' 'gory' 'claims' 'appeal' 'prosecution' 'claimed' 'poisoning' 'placed' 'stab' 'builds' 'narcissistic' 'scuttled' 'impossible' 'head' 'days' 'admitting' 'miles' 'rescued' 'broadcasting' 'buildup' 'began' 'dismissed' 'aloud' 'take' 'torturing' 'torso' 'hatches' 'small' 'fumes' 'without' 'engmark' 'news' 'alarm' 'psychiatrists' '70kg' 'especially' 'cited' 'would' 'describe' 'contributed' 'dismemberment' 'initially' 'assessment' 'repeatedly' 'normally' 'stood' 'told' 'wednesday' 'photograph' 'trip' 'gain' 'company' 'community' 'two' 'result' 'scanpix' 'died' 'claiming' 'serious' 'ship' 'copenhagen' 'insane' 'unable' 'others' 'incident' 'thursday' 'stay' 'listening' 'reportedly' 'murdering' 'divers' 'wired' 'cutting']
--	--	--

Ke-1	<p>['ms' '40' '6' 'tortured' 'plastic' 'formidable' 'may' 'driven' 'far' 'offer' 'london' 'following' 'unsupported' 'a board' 'next' 'return' 'freelance' '1987' 'emotionally' 'hald' 'anna' 'crimes' 'studied' 'oxbridge' 'south' 'raised' 'overboard' 'films' 'born' 'koge' 'ordered' 'undocumented' 'missing' 'crime' 'appear' 'personal' 'killed' 'prosecutors' '50km' 'media' 'carbon' 'semen' '30' 'case' 'woman' 'horrible' 'relations' 'manner' 'guilt' 'friends' 'weighted' 'offering' 'town' 'homemade' 'press' 'rado' 'bags' 'honours' 'trelleborg' 'motive' 'throat' 'character' 'tendencies' 'quoted' 'evidence' 'say' 'baltic' '10' 'conviction' '12' 'accepted' '8' 'knit' 'discovered' 'claim' 'bury' 'previously' 'interview' 'torture' 'school' 'venture' 'within' 'year' 'credible' 'dividing' 'journalism' 'terrible' 'october' 'detailed' 'happened' 'genitals' 'masters' 'arrest' 'bay' 'day' 'betina' 'set' 'consistent' 'telling' 'classmate' 'cyclist' 'top' 'close' 'economics' 'account' 'events' 'opening' 'university' 'board' 'interviewed' 'suffocate' 'unconvincing' 'researching' 'arguing' 'randomly' 'develops' 'police' 'departed' 'spacecraft' 'suspicion' 'bed' 'women' 'codrea' 'decision' 'accidental' 'featured' 'either' 'playback' 'go' 'vessel' 'timeline' 'somewhere' 'monoxide' 'clothing' '11' 'winning' 'sweden' 'assault' 'stabbed' 'device' 'computer' 'parole' '2017' 'acted' 'remains' 'conceal' 'differing' 'dismember' 'even' 'afp' 'month' 'gave' 'mutilated' 'sank' 'international' 'caused' 'session' 'programme' 'aug' 'underwear' 'waters' 'copyright' 'deliberately' 'spotted' 'kill' 'convict' 'suffocating' 'horror' 'charged' 'impaired' 'anger' 'helped' 'accounts'</p>	<p>['eastern' 'helmets' 'several' 'died' 'volunteer' 'bodies' 'rise' 'deaths' 'syria' 'rescuers' 'white' 'graphic' 'likely' 'force' 'showing' 'tweeted' 'basements' 'medics' 'last' 'ghouta']</p>
------	--	--

	<p>'made' 'impaled' 'upwards' 'sky' 'allegation' s' 'toxic' 'investigators' 'manned' 'old' 'columbia' 'associated' 'boar ding' 'since' 'co' 'dumped' 'southern' 'believe' 'class' 'across' 'know' 's howing' '22' '150lb' 'wounds' 'cohort' 'chosen' 'lift' 'strait' 'grew' 'stockings' '21' 'passing' 'based' 'opinion' 'founder' 'friend' 'weeks' 'tonne' 'claims' 'poisoning' 'placed' 'stab' 'builds' 'scuttled' ' went' 'admitting' 'miles' 'rescued' 'buildup' 'began' 'dismissed' 'smal l' 'fumes' 'without' 'engmark' 'alarm' '70kg' 'self' 'night' 'especi ally' 'describe' 'contributed' 'caption' 'gain' 'company' 'co mmunity' 'result' 'serious' 'insane' 'unable' 'incident' 'built' 'image' 'di vers']</p>	
Ke-2	<p>['many' 'least' 'international' 'apparent' '85' ' injured' 'statement' 'assad' 'shocking' 'douma' 'outrage' 'women' 'suspected' 'found' 'rescue' '70' 'children' 'people' 'big' 'price' 'medical' ' syrian' 'iran' 'left' 'rebel' 'local' 'lashed' 'workers' 'bashar' 'atta ck' 'trump' 'mostly' 'killed' 'sparked' 'following' 'held' 'fighting' 'suggest' 'evening' 'reports' 'difficulty' 'revulsion' 'allies' 'imag es' 'chemical' 'gas' 'world' 'according' 'al' 'including' 'lives' 'see n' 'president' 'saying' 'around' 'russia' 'widespread' 'pay' 'say' 'bea med' 'may' 'civilians' '500' 'said' 'town' 'shelters' 'families' 'staff' ' donald' 'us' 'saturday' 'sources' 'dozens' 'breathing' 'homes' 'centre s']</p>	<p>['coast' 'motivated' 'deserve' 'burko e' 'argued' 'inside' 'first' 'court' 'equates' 'empathy' 'story' 'city' 'jo urnalist' 'outside' 'also' 'jepsen' 'taken' 'ended' 'sentenced' 'given' ' danish' 'finally' 'activity' 'legs' 'necessary' 'divide' 'removed' 'deta ils' 'pressure' 'indecent' 'help' 'aid' 'defense' 'murder' 'kim' 'inclu ding' 'danger' 'clear' 'exhibited' 'images' 'ap' 'accident' 'corpse' 'ma dsen' 'jurors' 'hit' 'sought' 'various' '47' 'seen' 'written' 'metal' 'team' 'committed' 'read' 'extended' 'never' 'private' 'deaths' ' pleaded' 'peter' 'dropped' 'planned' 'severe' 'uc3' 'via' 'hours' 'legal' 'make' 'early' 'piece' 'sea' 'respect' 'hatch' 'wanted' 'beiji ng' 'added' 'established' 'explanations' 'multiple' 'sentence' 'explanation' 'fell' 'boyfriend' 'years' 'parts' 'wonderful' 'accordin g' 'photos' 'magazine' 'prior' 'premeditated' 'place' 'problem' 'la wyer' 'excerpts' 'something' 'moved' 'quietly' 'denied' 'verdict' 'handling' 'burying' 'new' 'publications'</p>

		'bax' 'later' '16' 'sexual' 'report' 're morse' 'said' 'showed' 'traits' 'region' 'admitted' 'brutal' 'describe d' 'family' 'lindhardt' 'sexually' 'decided' 'getty' 'whether' 'trial' 'ex plaining' 'island' 'along' 'throwing' 'took' 'spare' 'judge' 'sw edish' 'psychopathic' 'determined' 'emotion' 'agreed' 'confessed' 'nord ic' 'changed' 'prosecutor' 'several' 'reached' 'stopped' 'psychiatric' 'tes tifying' 'authorities' 'ride' 'tried' 'saying' 'found' 'restrictions' 'offered' 'buch' 'lifeless' 'march' 'conducted' 'dismembering' 'good' 'denmark' 'arms' 'sailing' 'trustworthy' 'lack' 'cause' 'body' 'p itch' 'sudden' 'pieces' 'accused' 'videos' 'open' 'nautilus' 'one' 'abro ad' 'immediately' 'walls' 'shows' 'heard' 'accidentally' 'disappeared' 'anette' 'prison' 'york' 'gruesome' 'death' 'special' 'life' 'smaller' 'last' 'hearing' 'shifting' 'guardian' 'departing' 'large' 'deck' 'tools' 'giv e' 'reuters' 'dead' 'risk' 'tied' 'guilty' 'submarine' 'evening' 'coul d' 'sink' 'jakob' 'times' 'dismembered' 'august' 'bbc' 'broug ht' 'big' 'gory' 'appeal' 'prosecution' 'claimed' 'narcissistic' 'impossible' 'head' 'days' 'broadcasting' 'aloud' 'take' 'torturing' 'torso' 'hatches' 'ne ws' 'psychiatrists' 'cited' 'would' 'inventor' 'dismemberment' 'initially' 'assessment' 'repeatedly' 'normally' 'stood' 'told' 'wednesday' 'photograph' 'trip' 'two' 'scanpix' 'died' 'claiming' 'ship' 'copenhagen' 'others' 'wall' 'thursday' 'stay' 'listening' 'reportedly' 'murdering' ' wired' 'cutting']
Ke-3	['ms' '40' '6' 'tortured' 'formidable' 'driven' ' far' 'london' 'unsupported' 'next' 'return' '1987' 'journali st' 'anna' 'studied' 'oxbridge' 'south' 'legs' 'raised' 'overboard' ' films' 'born' 'koge'	['coast' 'motivated' 'plastic' 'may' 'd eserve' 'burkoe' 'offer' 'argued' 'following' 'inside' 'aboard' 'first' 'c ourt' 'equates' 'empathy' 'story' 'freelance' 'city' 'emotionally' 'hald ' 'outside' 'also' 'jepsen' 'taken'

<p>'missing' 'prosecutors' '50km' 'media' 'carb on' '30' 'relations' 'seen'</p> <p>'friends' 'weighted' 'town' 'homemade' 'rad o' 'bags' 'honours'</p> <p>'trelleborg' 'throat' 'planned' 'character' 'ten dencies' 'say' '10' 'knit'</p> <p>'discovered' 'claim' 'previously' 'school' 've nture' 'within' 'year'</p> <p>'boyfriend' 'dividing' 'journalism' 'terrible'</p> <p>'october' 'happened'</p> <p>'masters' 'arrest' 'bay' 'day' 'telling' 'classm ate' 'cyclist' 'top'</p> <p>'close' 'economics' 'account' 'events' 'openi ng' 'university' 'board'</p> <p>'interviewed' 'unconvincing' 'researching' ' new' 'police' 'departed'</p> <p>'suspicion' 'women' 'codrea' 'accidental' 'de scribed' 'either' 'family'</p> <p>'getty' 'playback' 'vessel' 'somewhere' 'mon oxide' 'clothing' '11'</p> <p>'winning' 'sweden' 'assault' 'device' 'compu ter' 'parole' 'remains'</p> <p>'differing' 'even' 'afp' 'month' 'gave' 'mutila ted' 'sank' 'international'</p> <p>'session' 'programme' 'waters' 'copyright' ' deliberately' 'spotted' 'kill'</p> <p>'convict' 'suffocating' 'dismembering' 'help ed' 'accounts' 'investigators'</p> <p>'columbia' 'boarding' 'since' 'dumped' 'sout hern' 'believe' 'disappeared'</p> <p>'class' 'across' 'know' 'showing' '22' '150lb'</p> <p>'last' 'cohort' 'strait'</p> <p>'grew' '21' 'dismembered' 'passing' 'august'</p> <p>'bbc' 'friend' 'tonne'</p> <p>'poisoning' 'placed' 'scuttled' 'went' 'days' ' admitting' 'miles'</p> <p>'rescued' 'dismissed' 'small' 'without' 'alar m' '70kg' 'self' 'night'</p> <p>'describe' 'caption' 'gain' 'community' 'buil t' 'image' 'divers']</p>	<p>'ended' 'sentenced' 'crimes' 'given'</p> <p>'danish' 'finally' 'activity'</p> <p>'necessary' 'divide' 'removed' 'deta ils' 'pressure' 'indecent' 'ordered'</p> <p>'help' 'undocumented' 'crime' 'appe ar' 'personal' 'killed' 'aid' 'defense'</p> <p>'murder' 'kim' 'including' 'danger' ' clear' 'exhibited' 'images' 'semen'</p> <p>'ap' 'accident' 'corpse' 'case' 'madse n' 'jurors' 'hit' 'sought' 'woman'</p> <p>'horrible' 'various' '47' 'manner' 'gu ilt' 'written' 'metal' 'team'</p> <p>'offering' 'committed' 'read' 'exten ded' 'press' 'never' 'private'</p> <p>'deaths' 'pleaded' 'peter' 'motive' 'd ropped' 'quoted' 'severe' 'evidence'</p> <p>'uc3' 'via' 'baltic' 'hours' 'legal' 'ma ke' 'early' 'conviction' '12'</p> <p>'piece' 'sea' 'accepted' '8' 'respect' ' hatch' 'wanted' 'beijing' 'bury'</p> <p>'interview' 'torture' 'added' 'establi shed' 'explanations' 'multiple'</p> <p>'sentence' 'explanation' 'fell' 'credi ble' 'years' 'parts' 'wonderful'</p> <p>'according' 'photos' 'detailed' 'genit als' 'magazine' 'prior' 'betina'</p> <p>'premeditated' 'place' 'problem' 'set ' 'lawyer' 'excerpts' 'consistent'</p> <p>'something' 'moved' 'quietly' 'deni ed' 'verdict' 'handling' 'suffocate'</p> <p>'burying' 'publications' 'bax' 'later'</p> <p>'16' 'sexual' 'arguing' 'randomly'</p> <p>'report' 'develops' 'remorse' 'said' 's howed' 'spacecraft' 'traits'</p> <p>'region' 'bed' 'decision' 'admitted' ' brutal' 'featured' 'lindhardt'</p> <p>'sexually' 'decided' 'go' 'whether' 't rial' 'explaining' 'island' 'along'</p> <p>'throwing' 'took' 'spare' 'timeline' 'j udge' 'swedish' 'stabbed'</p> <p>'psychopathic' 'determined' '2017' ' acted' 'emotion' 'conceal' 'agreed'</p> <p>'confessed' 'nordic' 'changed' 'pros ecutor' 'dismember' 'several'</p> <p>'reached' 'caused' 'stopped' 'psychi atric' 'aug' 'underwear' 'testifying'</p> <p>'authorities' 'ride' 'tried' 'saying' 'fo und' 'horror' 'restrictions'</p>
--	---

		<p>'offered' 'charged' 'buch' 'lifeless' 'march' 'impaired' 'anger'</p> <p>'conducted' 'good' 'made' 'impaled' 'denmark' 'arms' 'upwards' 'sailing' 'trustworthy' 'lack' 'sky' 'cause' 'body' 'pitch' 'allegations' 'toxic'</p> <p>'manned' 'sudden' 'pieces' 'accused' 'old' 'videos' 'open' 'associated'</p> <p>'nautilus' 'one' 'abroad' 'co' 'immediately' 'walls' 'shows' 'heard'</p> <p>'accidentally' 'anette' 'prison' 'york' 'gruesome' 'death' 'special'</p> <p>'life' 'smaller' 'hearing' 'shifting' 'wounds' 'guardian' 'departing'</p> <p>'chosen' 'large' 'lift' 'deck' 'tools' 'give' 'reuters' 'dead' 'risk'</p> <p>'tied' 'stockings' 'guilty' 'submarine' 'evening' 'could' 'sink' 'jakob'</p> <p>'times' 'based' 'opinion' 'founder' 'weeks' 'brought' 'big' 'gory'</p> <p>'claims' 'appeal' 'prosecution' 'claimed' 'stab' 'builds' 'narcissistic'</p> <p>'impossible' 'head' 'broadcasting' 'buildup' 'began' 'aloud' 'take'</p> <p>'torturing' 'torso' 'hatches' 'fumes' 'engmark' 'news' 'psychiatrists'</p> <p>'especially' 'cited' 'would' 'inventor' 'contributed' 'dismemberment'</p> <p>'initially' 'assessment' 'repeatedly' 'normally' 'stood' 'told'</p> <p>'wednesday' 'photograph' 'trip' 'company' 'two' 'result' 'scanpix' 'died'</p> <p>'claiming' 'serious' 'ship' 'copenhagen' 'insane' 'unable' 'others' 'wall'</p> <p>'incident' 'thursday' 'stay' 'listening' 'reportedly' 'murdering' 'wired' 'cutting']</p>
Ke-4	<p>['ms' '40' '6' 'tortured' 'formidable' 'driven' 'far' 'london'</p> <p>'unsupported' 'next' 'return' '1987' 'journalist' 'also' 'anna'</p> <p>'sentenced' 'studied' 'oxbridge' 'south' 'legs' 'raised' 'overboard'</p> <p>'films' 'born' 'koge' 'missing' 'prosecutors' '50km' 'media' 'carbon'</p> <p>'images' '30' 'hit' 'relations' 'seen' 'friends' 'weighted' 'town'</p> <p>'homemade' 'rado' 'bags' 'honours' 'trelleborg' 'throat' 'planned'</p>	<p>['coast' 'motivated' 'plastic' 'may' 'deserve' 'burkoe' 'offer' 'argued'</p> <p>'following' 'inside' 'aboard' 'first' 'court' 'equates' 'empathy' 'story'</p> <p>'freelance' 'city' 'emotionally' 'held' 'outside' 'jepson' 'taken' 'ended'</p> <p>'crimes' 'given' 'danish' 'finally' 'activity' 'necessary' 'divide'</p> <p>'removed' 'details' 'pressure' 'indecent' 'ordered' 'help' 'undocumented'</p> <p>'crime' 'appear' 'personal' 'killed' 'aid' 'defense' 'murder' 'kim'</p>

<p>'character' 'tendencies' 'say' '10' 'knit' 'hatch' 'discovered' 'claim'</p> <p>'previously' 'school' 'venture' 'within' 'year' 'boyfriend' 'dividing'</p> <p>'journalism' 'terrible' 'october' 'happened' 'masters' 'arrest' 'bay'</p> <p>'day' 'premeditated' 'place' 'telling' 'classmate' 'cyclist' 'top' 'close'</p> <p>'economics' 'account' 'events' 'opening' 'university' 'board'</p> <p>'interviewed' 'unconvincing' 'researching' 'new' 'police' 'departed'</p> <p>'suspicion' 'women' 'codrea' 'admitted' 'accidental' 'described' 'either'</p> <p>'family' 'getty' 'playback' 'vessel' 'throwing' 'somewhere' 'monoxide'</p> <p>'clothing' '11' 'winning' 'sweden' 'swedish' 'assault' 'device' 'computer'</p> <p>'parole' 'remains' 'differing' 'changed' 'even' 'afp' 'month' 'gave'</p> <p>'mutilated' 'sank' 'international' 'session' 'programme' 'waters'</p> <p>'copyright' 'deliberately' 'spotted' 'kill' 'convict' 'suffocating'</p> <p>'dismembering' 'helped' 'accounts' 'investigators' 'columbia' 'boarding'</p> <p>'since' 'dumped' 'southern' 'believe' 'disappeared' 'class' 'across'</p> <p>'know' 'showing' '22' '150lb' 'last' 'shifting' 'cohort' 'strait' 'deck'</p> <p>'grew' 'submarine' 'evening' '21' 'dismembered' 'passing' 'august' 'bbc'</p> <p>'friend' 'tonne' 'poisoning' 'placed' 'scuttled' 'head' 'went' 'days'</p> <p>'admitting' 'miles' 'rescued' 'dismissed' 'small' 'without' 'alarm' '70kg'</p> <p>'self' 'night' 'describe' 'caption' 'gain' 'community' 'built' 'image'</p> <p>'divers']</p>	<p>'including' 'danger' 'clear' 'exhibited' 'semen' 'ap' 'accident' 'corpse'</p> <p>'case' 'madsen' 'jurors' 'sought' 'woman' 'horrible' 'various' '47'</p> <p>'manner' 'guilt' 'written' 'metal' 'team' 'offering' 'committed' 'read'</p> <p>'extended' 'press' 'never' 'private' 'deaths' 'pleaded' 'peter' 'motive'</p> <p>'dropped' 'quoted' 'severe' 'evidence' 'uc3' 'via' 'baltic' 'hours'</p> <p>'legal' 'make' 'early' 'conviction' '12' 'piece' 'sea' 'accepted' '8'</p> <p>'respect' 'wanted' 'beijing' 'bury' 'interview' 'torture' 'added'</p> <p>'established' 'explanations' 'multiple' 'sentence' 'explanation' 'fell'</p> <p>'credible' 'years' 'parts' 'wonderful' 'according' 'photos' 'detailed'</p> <p>'genitals' 'magazine' 'prior' 'betina' 'problem' 'set' 'lawyer' 'excerpts'</p> <p>'consistent' 'something' 'moved' 'quietly' 'denied' 'verdict' 'handling'</p> <p>'suffocate' 'burying' 'publications' 'bax' 'later' '16' 'sexual' 'arguing'</p> <p>'randomly' 'report' 'develops' 'remorse' 'said' 'showed' 'spacecraft'</p> <p>'traits' 'region' 'bed' 'decision' 'brutal' 'featured' 'lindhardt'</p> <p>'sexually' 'decided' 'go' 'whether' 'trial' 'explaining' 'island' 'along'</p> <p>'took' 'spare' 'timeline' 'judge' 'stabbed' 'psychopathic' 'determined'</p> <p>'2017' 'acted' 'emotion' 'conceal' 'agreed' 'confessed' 'nordic'</p> <p>'prosecutor' 'dismember' 'several' 'reached' 'caused' 'stopped'</p> <p>'psychiatric' 'aug' 'underwear' 'testifying' 'authorities' 'ride' 'tried'</p> <p>'saying' 'found' 'horror' 'restriction' 's' 'offered' 'charged' 'buch'</p> <p>'lifeless' 'march' 'impaired' 'anger' 'conducted' 'good' 'made' 'impaled'</p> <p>'denmark' 'arms' 'upwards' 'sailing' 'trustworthy' 'lack' 'sky' 'cause'</p> <p>'body' 'pitch' 'allegations' 'toxic' 'manned' 'sudden' 'pieces' 'accused'</p> <p>'old' 'videos' 'open' 'associated' 'nautilus' 'one' 'abroad' 'co'</p>
---	---

		<p>'immediately' 'walls' 'shows' 'hear d' 'accidentally' 'anette' 'prison' 'york' 'gruesome' 'death' 'special' 'life' 'smaller' 'hearing' 'wounds' 'guardian' 'departing' 'chosen' 'large' 'lift' 'tools' 'give' 'reuters' 'dead' 'risk' 'tied' 'stockings' 'guilty' 'could' 'sink' 'jakob' 'times' 'based' 'opinion' 'founder' 'weeks' 'brought' 'big' 'gory' 'claims' 'appeal' 'prosecution' 'claimed' 'stable' 'builds' 'narcissistic' 'impossible' 'broadcasting' 'buildup' 'began' 'aloud' 'take' 'torturing' 'torso' 'hatches' 'fumes' 'engmark' 'news' 'psychiatrists' 'especially' 'cited' 'would' 'inventor' 'contributed' 'dismemberment' 'initially' 'assessment' 'repeatedly' 'normally' 'stood' 'told' 'wednesday' 'photograph' 'trip' 'company' 'two' 'result' 'scanpix' 'died' 'claiming' 'serious' 'ship' 'copenhagen' 'insane' 'unable' 'others' 'wall' 'incident' 'thursday' 'stay' 'listening' 'reported' 'murdering' 'wired' 'cutting']</p>
Ke-5	<p>['ms' '40' '6' 'tortured' 'formidable' 'driven' 'far' 'london' 'unsupported' 'next' 'story' 'return' '1987' 'journalist' 'also' 'anna' 'sentenced' 'studied' 'oxbridge' 'danish' 'south' 'legs' 'raised' 'overboard' 'films' 'born' 'koge' 'missing' 'prosecutors' '50km' 'kim' 'media' 'carbon' 'images' '30' 'hit' '47' 'relations' 'seen' 'friends' 'weighted' 'town' 'homemade' 'radio' 'bags' 'honours' 'trelleborg' 'dropped' 'throat' 'planned' 'character' 'tendencies' 'say' '10' 'sea' 'knit' 'hatch' 'discovered' 'claim' 'previously' 'school' 'venture' 'within' 'year' 'boyfriend' 'dividing' 'journalism' 'terrible' 'october' 'happened' 'masters' 'arrest' 'bay' 'day' 'premeditated' 'place' 'telling' 'classmate' 'cyclist' 'top' 'close' 'economics' 'account' 'events'</p>	<p>['coast' 'motivated' 'plastic' 'may' 'deserve' 'burkoe' 'offer' 'argued' 'following' 'inside' 'aboard' 'first' 'court' 'equates' 'empathy' 'freelance' 'city' 'emotionally' 'hald' 'outside' 'jepsen' 'taken' 'ended' 'crimes' 'given' 'finally' 'activity' 'necessary' 'divide' 'removed' 'details' 'pressure' 'indecent' 'ordered' 'help' 'undocumented' 'crime' 'appear' 'personal' 'killed' 'aid' 'defense' 'murder' 'including' 'danger' 'clear' 'exhibited' 'semen' 'ap' 'accident' 'corpse' 'case' 'madsen' 'jurors' 'sought' 'woman' 'horrible' 'various' 'manner' 'guilt' 'written' 'metal' 'team' 'offering' 'committed' 'read' 'extended' 'press' 'never' 'private' 'deaths' 'pleaded' 'peter' 'motive' 'quoted' 'severe' 'evidence' 'uc3' 'via' 'baltic' 'hours' 'legal' 'make' 'early' 'conviction' '12'</p>

	<p>'opening' 'university' 'board' 'interviewed' 'unconvincing' 'researching'</p> <p>'new' 'police' 'departed' 'suspicion' 'women' 'codrea' 'admitted'</p> <p>'accidental' 'described' 'either' 'family' 'getty' 'playback' 'vessel'</p> <p>'throwing' 'somewhere' 'monoxide' 'clothing' '11' 'winning' 'sweden'</p> <p>'swedish' 'assault' 'psychopathic' 'device' 'computer' 'parole' 'remains'</p> <p>'differing' 'changed' 'even' 'afp' 'month' 'gave' 'mutilated' 'sank'</p> <p>'international' 'session' 'programme' 'waters' 'copyright' 'deliberately'</p> <p>'spotted' 'kill' 'convict' 'suffocating' 'dismembering' 'helped'</p> <p>'accounts' 'body' 'investigators' 'columbia' 'boarding' 'since' 'dumped'</p> <p>'southern' 'believe' 'disappeared' 'class' 'across' 'know' 'showing' '22'</p> <p>'150lb' 'last' 'shifting' 'cohort' 'strait' 'deck' 'grew' 'submarine'</p> <p>'evening' '21' 'dismembered' 'passing' 'august' 'bbc' 'friend' 'tonne'</p> <p>'poisoning' 'placed' 'scuttled' 'head' 'went' 'days' 'admitting' 'miles'</p> <p>'rescued' 'dismissed' 'small' 'without' 'alarm' '70kg' 'self' 'night'</p> <p>'describe' 'initially' 'caption' 'gain' 'community' 'built' 'image'</p> <p>'divers']</p>	<p>'piece' 'accepted' '8' 'respect' 'wanted' 'beijing' 'bury' 'interview'</p> <p>'torture' 'added' 'established' 'explanations' 'multiple' 'sentence'</p> <p>'explanation' 'fell' 'credible' 'years' 'parts' 'wonderful' 'according'</p> <p>'photos' 'detailed' 'genitals' 'magazine' 'prior' 'betina' 'problem' 'set'</p> <p>'lawyer' 'excerpts' 'consistent' 'something' 'moved' 'quietly' 'denied'</p> <p>'verdict' 'handling' 'suffocate' 'burying' 'publications' 'bax' 'later'</p> <p>'16' 'sexual' 'arguing' 'randomly' 'report' 'develops' 'remorse' 'said'</p> <p>'showed' 'spacecraft' 'traits' 'region' 'bed' 'decision' 'brutal'</p> <p>'featured' 'lindhardt' 'sexually' 'decided' 'go' 'whether' 'trial'</p> <p>'explaining' 'island' 'along' 'took' 'spare' 'timeline' 'judge' 'stabbed'</p> <p>'determined' '2017' 'acted' 'emotion' 'conceal' 'agreed' 'confessed'</p> <p>'nordic' 'prosecutor' 'dismember' 'several' 'reached' 'caused' 'stopped'</p> <p>'psychiatric' 'aug' 'underwear' 'testifying' 'authorities' 'ride' 'tried'</p> <p>'saying' 'found' 'horror' 'restriction' 's' 'offered' 'charged' 'buch'</p> <p>'lifeless' 'march' 'impaired' 'anger' 'conducted' 'good' 'made' 'impaled'</p> <p>'denmark' 'arms' 'upwards' 'sailing' 'trustworthy' 'lack' 'sky' 'cause'</p> <p>'pitch' 'allegations' 'toxic' 'manned' 'sudden' 'pieces' 'accused' 'old'</p> <p>'videos' 'open' 'associated' 'nautilus' 'one' 'abroad' 'co' 'immediately'</p> <p>'walls' 'shows' 'heard' 'accidentally' 'anette' 'prison' 'york' 'gruesome'</p> <p>'death' 'special' 'life' 'smaller' 'hearing' 'wounds' 'guardian'</p> <p>'departing' 'chosen' 'large' 'lift' 'tools' 'give' 'reuters' 'dead' 'risk'</p> <p>'tied' 'stockings' 'guilty' 'could' 'sink' 'jakob' 'times' 'based'</p> <p>'opinion' 'founder' 'weeks' 'brought' 'big' 'gory' 'claims' 'appeal'</p> <p>'prosecution' 'claimed' 'stab' 'build' 's' 'narcissistic' 'impossible'</p>
--	--	---

		'broadcasting' 'buildup' 'began' 'aloud' 'take' 'torturing' 'torso' 'hatches' 'fumes' 'engmark' 'news' 'psychiatrists' 'especially' 'cited' 'would' 'inventor' 'contributed' 'dismemberment' 'assessment' 'repeatedly' 'normally' 'stood' 'told' 'wednesday' 'photograph' 'trip' 'company' 'two' 'result' 'scanpix' 'died' 'claiming' 'serious' 'ship' 'copenhagen' 'insane' 'unable' 'others' 'wall' 'incident' 'thursday' 'stay' 'listening' 'reportedly' 'murdering' 'wired' 'cutting']
Ke-6	['ms' '40' '6' 'tortured' 'formidable' 'driven' 'far' 'london' 'unsupported' 'next' 'story' 'return' '1987' 'journalist' 'also' 'anna' 'sentenced' 'studied' 'oxbridge' 'danish' 'south' 'legs' 'raised' 'overboard' 'films' 'born' 'koge' 'missing' 'prosecutors' '50km' 'murder' 'kim' 'media' 'carbon' 'images' '30' 'hit' '47' 'relations' 'seen' 'friends' 'weighted' 'town' 'homemade' 'radio' 'bags' 'honours' 'trelleborg' 'dropped' 'throat' 'planned' 'character' 'tendencies' 'say' '10' 'sea' 'knit' 'hatch' 'discovered' 'claim' 'previously' 'school' 'venture' 'within' 'year' 'boyfriend' 'dividing' 'journalism' 'terrible' 'october' 'happened' 'masters' 'arrest' 'bay' 'day' 'premeditated' 'place' 'telling' 'classmate' 'cyclist' 'top' 'close' 'economics' 'account' 'events' 'opening' 'university' 'board' 'interviewed' 'unconvincing' 'researching' 'new' 'police' 'departed' 'suspicion' 'women' 'codrea' 'admitted' 'accidental' 'described' 'either' 'family' 'getty' 'playback' 'vessel' 'throwing' 'somewhere' 'monoxide' 'clothing' '11' 'winning' 'sweden' 'swedish' 'assault' 'psychopathic' 'device' 'computer' 'parole' 'remains' 'differing' 'changed' 'even' 'afp' 'month' 'gave' 'mutilated'	['coast' 'motivated' 'plastic' 'may' 'deserve' 'burkoe' 'offer' 'argued' 'following' 'inside' 'aboard' 'first' 'court' 'equates' 'empathy' 'freelance' 'city' 'emotionally' 'held' 'outside' 'jepsen' 'taken' 'ended' 'crimes' 'given' 'finally' 'activity' 'necessary' 'divide' 'removed' 'details' 'pressure' 'indecent' 'ordered' 'help' 'undocumented' 'crime' 'appear' 'personal' 'killed' 'aid' 'defense' 'including' 'danger' 'clear' 'exhibited' 'semen' 'ap' 'accident' 'corpse' 'case' 'madsen' 'jurors' 'sought' 'woman' 'horrible' 'various' 'manner' 'guilt' 'written' 'metal' 'team' 'offering' 'committed' 'read' 'extended' 'press' 'never' 'private' 'deaths' 'pleaded' 'peter' 'motive' 'quoted' 'severe' 'evidence' 'uc3' 'via' 'baltic' 'hours' 'legal' 'make' 'early' 'conviction' '12' 'piece' 'accepted' '8' 'respect' 'wanted' 'beijing' 'bury' 'interview' 'torture' 'added' 'established' 'explanations' 'multiple' 'sentence' 'explanation' 'fell' 'credible' 'years' 'parts' 'wonderful' 'according' 'photos' 'detailed' 'genitals' 'magazine' 'prior' 'betina' 'problem' 'set' 'lawyer' 'excerpts' 'consistent' 'something' 'moved' 'quietly' 'denied' 'verdict' 'handling' 'suffocate' 'burying' 'publications' 'bax' 'later' '16'

	<p>'sank' 'international' 'session' 'programme' 'waters' 'copyright' 'deliberately' 'spotted' 'kill' 'convict' 'suffocating' 'dismembering' 'helped' 'accounts' 'body' 'investigators' 'columbia' 'boarding' 'since' 'dumped' 'southern' 'believe' 'disappeared' 'class' 'across' 'know' 'showing' '22' '150lb' 'last' 'shifting' 'cohort' 'strait' 'deck' 'grew' 'submarine' 'evening' '21' 'dismembered' 'passing' 'august' 'bbc' 'friend' 'tonne' 'poisoning' 'placed' 'scuttled' 'head' 'went' 'days' 'admitting' 'miles' 'rescued' 'dismissed' 'small' 'without' 'alarm' '70kg' 'self' 'night' 'describe' 'initially' 'caption' 'gain' 'community' 'copenhagen' 'wall' 'built' 'image' 'divers']</p>	<p>'sexual' 'arguing' 'randomly' 'report' 'develops' 'remorse' 'said' 'showed' 'spacecraft' 'traits' 'region' 'bed' 'decision' 'brutal' 'featured' 'lindhardt' 'sexually' 'decided' 'go' 'whether' 'trial' 'explaining' 'island' 'along' 'took' 'spare' 'timeline' 'judge' 'stabbed' 'determined' '2017' 'acted' 'emotion' 'conceal' 'agreed' 'confessed' 'nordic' 'prosecutor' 'dismember' 'several' 'reached' 'caused' 'stopped' 'psychiatric' 'aug' 'underwear' 'testifying' 'authorities' 'ride' 'tried' 'saying' 'found' 'horror' 'restriction' 'offered' 'charged' 'buck' 'lifeless' 'march' 'impaired' 'anger' 'conducted' 'good' 'made' 'impaled' 'denmark' 'arms' 'upwards' 'sailing' 'trustworthy' 'lack' 'sky' 'cause' 'pitch' 'allegations' 'toxic' 'manned' 'sudden' 'pieces' 'accused' 'old' 'videos' 'open' 'associated' 'nautilus' 'one' 'abroad' 'co' 'immediately' 'walls' 'shows' 'heard' 'accidentally' 'anette' 'prison' 'york' 'gruesome' 'death' 'special' 'life' 'smaller' 'hearing' 'wounds' 'guardian' 'departing' 'chosen' 'large' 'lift' 'tools' 'give' 'reuters' 'dead' 'risk' 'tied' 'stockings' 'guilty' 'could' 'sink' 'jakob' 'times' 'based' 'opinion' 'founder' 'weeks' 'brought' 'big' 'gory' 'claims' 'appeal' 'prosecution' 'claimed' 'stab' 'build' 's' 'narcissistic' 'impossible' 'broadcasting' 'buildup' 'began' 'aloud' 'take' 'torturing' 'torso' 'hatches' 'fumes' 'engmark' 'news' 'psychiatrists' 'especially' 'cited' 'would' 'inventor' 'contributed' 'dismemberment' 'assessment' 'repeatedly' 'normally' 'stood' 'told' 'wednesday' 'photograph' 'trip' 'company' 'two' 'result' 'scanpix' 'died' 'claiming' 'serious' 'ship' 'insane' 'unable' 'others' 'incident' 'thursday' 'stay' 'listening' 'reportedly' 'murdering'</p>
--	---	--

		'wired' 'cutting']
--	--	--------------------

TABEL D

Iterasi	Titik Tengah Cluster 1 (a)	Titik Tengah Cluster 2 (b)
Ke-0	[0.000000 0.000000 0.200000 0.000000]	[0.200000 0.000000 0.000000 0.200000]
Ke-1	[0.00151365 0.00141937 0. 00096681 0.00261151]	[0.00285215 0.00300002 0. 00353769 0.00214247]
Ke-2	[0.00156703 0.00154477 0.00130582 0.00216043]	[0.00181115 0.00185085 0.00190309 0.00157861]
Ke-3	[0.00163966 0.0016626 0.00151334 0.00197509]	[0.00172951 0.00172806 0.00167094 0.00171626]
Ke-4	[0.00166975 0.00169239 0.00157267 0.00189878]	[0.00170669 0.00170616 0.00161725 0.0017946]
Ke-5	[0.00168293 0.00170041 0.00159046 0.00186912]	[0.00169845 0.00170323 0.00160408 0.00182881]
Ke-6	[0.00168855 0.00170264 0.00 159605 0.00185776]	[0.00169499 0.00170319 0.00160053 0.00184241]
Ke-7	[0.00169087 0.00170329 0.00159791 0.00185344]	[0.0016935 0.00170339 0.00159946 0.00184763]
Ke-8	[0.00169181 0.00170349 0.00159855 0.0018518]	[0.00169287 0.0017035 0.00159911 0.0018496]
Ke-9	[0.00169219 0.00170355 0.00159878 0.00185118]	[0.0016926 0.00170355 0.00159899 0.00185035]
Ke-10	[0.00169234 0.00170357 0.00159886 0.00185094]	[0.0016925 0.00170357 0.00159894 0.00185063]

BIODATA PENULIS



Penulis bernama lengkap **Adzra Nazihah**, lahir di Jakarta, 27 Juni 1996. Anak pertama dari pasangan Mardanus dan Eva Nur Efrah, serta memiliki dua adik laki-laki bernama Abdul Fattah Afif Ismail dan Muhammad Emir Musyaffa. Penulis menempuh pendidikan di SDIT At-Taufiq Depok, SMPIT Nurul Fikri Depok, dan SMAN 1 Depok. Selanjutnya, penulis melanjutkan pendidikan tingginya di Departemen Matematika Institut Teknologi Sepuluh

Nopember (ITS) Surabaya dengan mengambil bidang minat Ilmu Komputer. Selama mengikuti perkuliahan di ITS, penulis turut aktif dalam beberapa kegiatan kemahasiswaan seperti menjadi Panitia Konferensi YES SUMMIT yang diadakan oleh BEM ITS, Panitia Pasar Malam Minggu ITS yang diadakan oleh BEM ITS, Sekretaris 1 Olimpiade Matematika ITS (OMITS), mengikuti pertukaran pelajar ke Dankook University di Korea Selatan pada tahun 2016, dll. Informasi lebih lanjut mengenai Tugas Akhir ini dapat ditujukan ke penulis melalui email: adzranzh@gmail.com.